APPENDIX A

COPIES OF RELEVANT REFERENCES IDENTIFIED IN DR. SAMPSON'S SEARCH

Oligonucleotides Containing 2-Aminoadenine and 2-Thiothymine Act as Selectively Binding Complementary Agents

Igor V. Kutyavin, Rebecca L. Rhinehart, Eugeny A. Lukhtanov, Vladimir V. Gorn, Rich B. Meyer, Jr., and Howard B. Gamper, Jr.*

Epoch Pharmaceuticals, Inc., 1725 220th Street SE, #104, Bothell, Washington 98021

Received March 13, 1996; Revised Manuscript Received June 11, 1996®

ABSTRACT: A pair of complementary oligodeoxynucleotides (ODNs) uniformly substituted with 2-aminoadenine (A′) in place of adenine and 2-thiothymine (T′) in place of thymine did not hybridize to each other but did form very stable hybrids with unmodified complementary ODNs. These unusual properties were a consequence of the hydrogen-bonding properties of the two base analogs. Thermal denaturation studies of short duplexes which contained these bases demonstrated that the A′-T and A-T′ doublets formed stable base pairs whereas the A′-T′ doublet acted like a mismatch. Complementary ODNs substituted with these base analogs are referred to as SBC or selectively binding complementary ODNs. When used as a pair, these single-stranded ODNs invaded the ends of homologous duplexes and formed stable three-arm junctions under conditions where unmodified ODNs failed to give a product. SBC ODNs have a fundamental thermodynamic advantage in hybridizing to short segments of double-stranded nucleic acid and represent a new approach for the design of oligomeric probes and antisense agents. Many secondary structure features present in long single-stranded nucleic acids should be accessible to these reagents.

The use of oligodeoxynucleotides (ODNs)[1] as diagnostic probes and antisense agents is based upon the Watson−Crick base pairing of complementary nucleic acid sequences. These same hydrogen-bonding interactions also occur in long single-stranded DNA or RNA molecules, where they result in the formation of short, usually imperfect duplexes which can interfere with the hybridization of ODNs (Gamper et al., 1987; Chastain & Tinoco, 1993). Although numerous approaches have been described for the design of ODNs which overcome or exploit the natural tendency of DNA or RNA to base pair with itself, none provides a general solution to the problem.

Modified ODNs which form very stable hybrids have frequently been used to improve the efficiency of hybridization to single-stranded nucleic acid. Examples include ODNs with 2′-modified (Monia et al., 1993; Sproat & Lamond, 1993), N3′ → P5′ phosphoramidate (Gryaznov & Chen, 1994), and peptide (Nielsen et al., 1994) backbones, ODNs containing base analogs such as 2-aminoadenine (Lamm et al., 1991) or C5 propynylpyrimidines (Wagner et al., 1993), or ODNs conjugated to an intercalating agent (Asseline et al., 1984) or a minor groove binding agent (Afonina et al., 1996). However, none of these modifications guarantees efficient hybridization if the targeted sequence is already substantially base paired to another sequence in the same molecule.

Hybridization strategies which rationally exploit specific sequence or structural elements within a single-stranded target have also been described. Circular ODNs capable of triple-strand formation with a homopurine or homopyrimidine run exhibit enhanced binding affinities relative to control ODNs which only form a duplex (Prakash & Kool, 1991; Wang & Kool, 1994). Tethered ODNs complementary to two single-stranded sequences in close proximity to one another (Richardson & Schepartz, 1991) or separate ODNs which bind to contiguous sequences (Kutyavin et al., 1988) can exhibit cooperative binding. Localized hairpins are frequently found in long single-stranded nucleic acids and can be rationally targeted. Formation of a pseudoknot by hybridization of an ODN to a hairpin loop can significantly enhance hybrid stability (Ecker et al., 1992; Lima et al., 1992). Alternatively, an ODN can hybridize to the two single-stranded arms of a hairpin (Francois et al., 1994). In the special case where the hairpin stem contains a homopurine run, a single ODN can bind to both the stem and one of the flanking single-stranded arms (Brossalina & Toulme, 1993; Francois & Helene, 1995).

We describe the hybridization and strand invasion properties of a new class of modified ODNs designed to be used as paired complements. Substituted with 2-thiothymine and 2-aminoadenine, these selectively binding complementary (SBC) ODNs form very stable hybrids with complementary, unmodified sequences, yet they do not interact with each other. Hybridization of a complementary pair of single-stranded SBC ODNs to both strands of a short homologous DNA or RNA duplex should be kinetically favored since the pathway does not involve the formation of an intermediate Holliday junction and thermodynamically favored due to an increase in the number of base pairs (see Figure 1). As a consequence, we show that two SBC ODNs (but not the corresponding unmodified ODNs) strand invade both blunt-ended and recessed duplexes to form stable three-arm junctions. These model experiments suggest that the hybridization of paired SBC ODNs to long single-stranded

* Corresponding author. Telephone: 206-485-8566. Fax: 206-486-8336. E-mail: howardg@epochpharm.com.

® Abstract published in *Advance ACS Abstracts*, August 1, 1996.

[1] Abbreviations: A′, 2-aminoadenine; CPG, controlled pore glass; HPLC, high-performance liquid chromatography; ODN, oligodeoxynucleotide; SBC, selectively binding complementary; TAJ, three-arm junction; T_m, melting temperature; T′, 2-thiothymine.

0006-2960/96/0435-11470$12.00/0

Selectively l

A

60 bp

B

40 bp

FIGURE 1: (A migration to yi of strands occ $MgCl_2$, the si hundreds of n double-strande helices (Lilley preserves total longer homolo; SBC ODN is the crossover single-strandec of the highly st by the formatic is indicated for each junction l

DNA or RN/ secondary str

MATERIAL

5′-*O*-(Dime ethyl)-*N*,*N*′-d the procedur Diaminopurin scribed by Fat bis(phenoxyac (2-cyanoethyl) essentially as al., 1991). A Support, Phar to the procedu 1993).

Oligonuclec on a Pharmaci using hexanol from the solid ammonia at phenoxyacetyl removed with (1:5:5, v/v/v) trityl-on ODN: 305 mm) reve to 40% aceti detritylation v temperature, t 2% solution of

Enzymatic L modified base: the purity of s was digested esterase I, DNa sates were anal × 150 mm) c

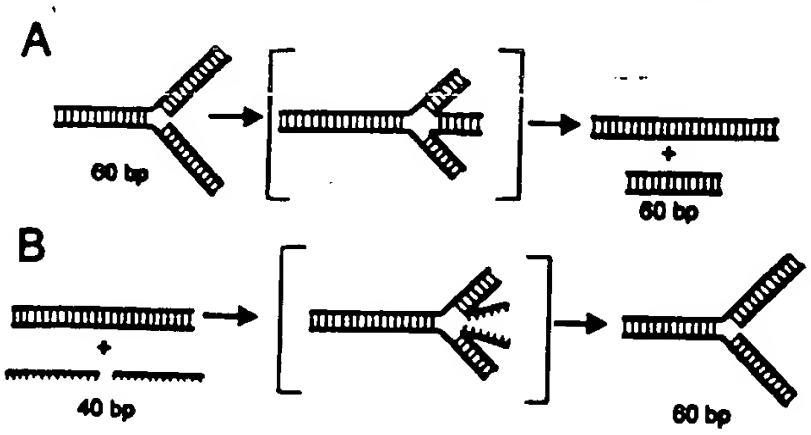


FIGURE 1: (A) Unmodified three-arm junction undergoes branch migration to yield two duplexes. During branch migration, crossover of strands occurs within a Holliday junction. In the presence of $MgCl_2$, the step time for branch migration is relatively long [hundreds of milliseconds (Panyutin et al., 1995)] because the four double-stranded arms of the junction stack to form two colinear helices (Lilley & Clegg, 1993). Branch migration in either direction preserves total base pairing. (B) Two SBC ODNs strand invade a longer homologous duplex to form a stable three-arm junction. Each SBC ODN is complementary to one strand of the duplex. Since the crossover junction contains three double-stranded and two single-stranded arms, its movement should be more facile than that of the highly structured Holliday junction. Strand invasion is driven by the formation of new base pairs. The total number of base pairs is indicated for the starting and ending states of the DNA, assuming each junction has three 20 bp long arms.

DNA or RNA should be less inhibited by the presence of secondary structure than a regular ODN.

MATERIALS AND METHODS

5′-*O*-(Dimethoxytrityl)-2-thiothymidine−3′-*O*-(2-cyanoethyl)-*N*,*N*′-diisopropylphosphoramidite was prepared using the procedure of Connolly and Newman (1989). 2,6-Diaminopurine 2′-deoxyriboside was synthesized as described by Fathi et al. (1990). 5′-*O*-(Dimethoxytrityl)-N^2,N^6-bis(phenoxyacetyl)-2,6-diaminopurine deoxyriboside−3′-*O*-(2-cyanoethyl)-*N*,*N*-diisopropylphosphoramidite was prepared essentially as described for the 2′-*O*-allyl analog (Sproat et al., 1991). A polystyrene support for DNA synthesis (Primer Support, Pharmacia) was modified with hexanol according to the procedure described for hexanol CPG (Gamper et al., 1993).

Oligonucleotide Synthesis. SBC ODNs were synthesized on a Pharmacia OligoPilot DNA synthesizer in 10 μmol scale using hexanol Primer Support. Deprotection and detachment from the solid support was accomplished with concentrated ammonia at 40 °C for 15 h. The remaining second phenoxyacetyl group on 2,6-diaminopurine residues was removed with a mixture of hydrazine/ethanolamine/methanol (1:5:5, v/v/v) (Polushin & Cohen, 1994). Purification of trityl-on ODNs was performed on a Hamilton PRP-1 (7.0 × 305 mm) reverse phase column employing a gradient of 5 to 40% acetonitrile in 0.1 M $NaClO_4$ (pH 7). After detritylation with 80% acetic acid for 15 min at room temperature, the ODNs were precipitated by addition of a 2% solution of $NaClO_4$ in acetone.

Enzymatic Digestion. To determine the stability of the modified bases to the deprotection conditions and evaluate the purity of synthesized ODNs, about 20 μg of each ODN was digested to nucleosides by a mixture of phosphodiesterase I, DNase I, and alkaline phosphatase. The hydrolysates were analyzed by reverse phase HPLC using a C18 (2 × 150 mm) column and a Waters 994 photodiode array detector. We detected less than 5% of the impurities derived from the modification of 2-thiothymine and 2,6-diaminopurine bases.

Thermal Denaturation Studies. Hybrids formed between complementary ODNs were melted at a rate of 0.5 °C/min in 200 mM NaCl, 0.1 mM EDTA, and 10 mM Na_2HPO_4 (pH 7.0) in a Lambda 2 (Perkin-Elmer) spectrophotometer equipped with a PTP-6 automatic multicell temperature programmer. Each ODN was mixed with an equimolar amount of complement to give a total strand concentration of 8×10^{-7} M. Prior to melting, samples were denatured at 100 °C and then cooled to the starting temperature over a 10 min period. The melting temperatures (T_m values) of the hybrids were determined from the derivative maxima. Free energies were calculated according to a "two-state" model by minimization of mean square errors between the calculated and experimental melting curves (Petershiem & Turner, 1983).

Gel Mobility Shift Assays. Sequential hybridization and strand invasion experiments were conducted at room temperature (25−27 °C) in 200 mM NaCl, 0.1 mM EDTA, and 10 mM Na_2HPO_4 (pH 7.0) as described in the figure legends. The concentrations of invading 20-mers and target duplex strands were 2.5×10^{-6} and 5×10^{-7} M, respectively. Prior to use, ODN 3 was 5′ end-labeled with ^{32}P using polynucleotide kinase and [γ-^{32}P]ATP. The final reaction volumes were 20−200 μL. Aliquots (5 μL) were removed at specific times into 5 μL volumes of cold dyes, ficoll and 2 mM $MgCl_2$, quickly frozen in a dry ice bath, and stored at −20 °C until they were analyzed. Immediately prior to loading onto a gel, each aliquot was thawed in an ice bath. Electrophoresis was conducted in a precooled 8% nondenaturing polyacrylamide gel (0.04 × 20 × 40 cm) containing 89 mM Tris-borate (pH 8.3), 2 mM EDTA, and 3 mM $MgCl_2$ for 4 h at 10 W. The dried gel was visualized by autoradiography, and bands were quantified using a BioRad GS-250 phosphorimager. Control studies showed that storage of aliquots prior to electrophoretic analysis did not alter the distribution of products.

RESULTS

Design of SBC ODNs. Selectively binding complementary ODNs are defined as self-complementary ODNs or pairs of complementary ODNs which do not interact with each other under physiological conditions and so exist as single-stranded molecules. While unable to base pair to each other, every SBC ODN can by design hybridize to a complementary DNA or RNA sequence with no mismatching. The hybrids so formed exhibit stabilities comparable to or greater than those of regular hybrids. These unique properties should promote strand invasion of paired SBC ODNs into a segment of double-stranded DNA or RNA. The kinetics of the reaction should be accelerated due to the reduced likelihood of forming transitory Holliday junctions, while the free energy of the complex should be reduced by the creation of additional base pairs. Figure 1 shows the stable three-arm junction formed when using SBC ODNs to strand invade a duplex. Also shown is the spontaneous decay of the corresponding unmodifed three-arm junction. These reactions will be described more fully once the experimental results have been presented.

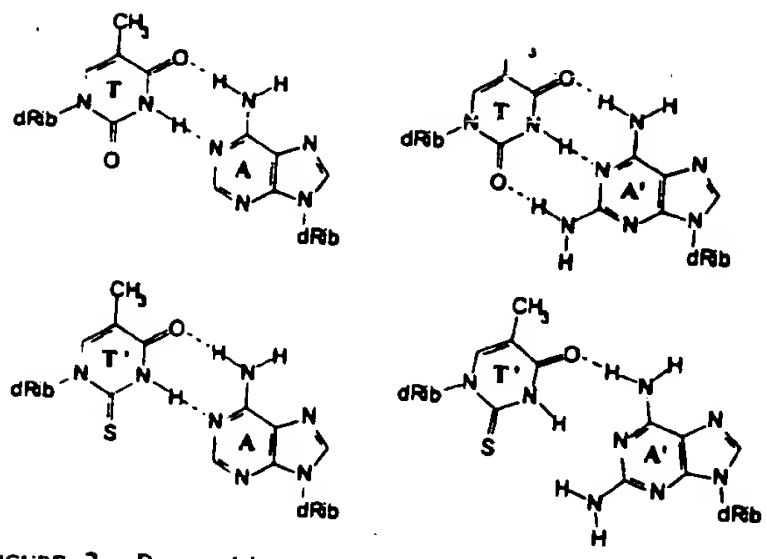


FIGURE 2: Base-pairing schemes for Watson–Crick doublets between adenine or 2-aminoadenine and thymine or 2-thiothymine.

Complementary 20-mers

ODN	Bases	Sequence
1a	A and T	5′-GTAAGAGAATTATGCAGTGC
1b	A′and T′	5′-GT′A′A′GA′GA′A′T′T′A′T′GCA′GT′GC
1c	A′and T	5′-GTA′A′GA′GA′A′TTA′TGCA′GTGC
1d	A and T′	5′-GT′AAGAGAAT′T′AT′GCAGT′GC
2a	A and T	5′-GCACTGCATAATTCTCTTAC
2b	A′and T′	5′-GCA′CT′GCA′T′A′A′T′T′CT′CT′T′A′C
2c	A′and T	5′-GCA′CTGCA′TA′A′TTCTCTTA′C
2d	A and T′	5′-GCACT′GCAT′AAT′T′CT′CT′T′AC

Double-Stranded 40-mers

ODN	Sequence
3	3′-ATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATG
4	5′-TATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTAC
3	3′-ATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATG
5	5′-TATCACTCATGGTTATGGCAGCACTGCATAATTCT

FIGURE 3: Oligonucleotides used in this study. **1a** and **2a** are normal (unmodified) ODNs, while **1b** and **2b** are SBC ODNs. The bold portions of hybrids **3**–**4** and **3**–**5** are homologous to ODNs **1a** and **2a**. All of the 20-mer ODNs were synthesized on a hexanol Primer Support, leading to the introduction of a 3′ hexanol phosphate cap. Use of a modified polystyrene support permitted the synthesis of SBC ODNs with 3′ terminal A′ or T′ bases.

To accomplish the design goals, SBC ODNs were synthesized by substituting 2-aminoadenine (A′) for adenine (A) and 2-thiothymine (T′) for thymine (T). Figure 2 shows the hydrogen-bonded pairs formed by these bases. The A′-T base pair possesses an extra hydrogen bond relative to A-T (Howard & Miles, 1984) and is frequently used to stabilize nucleic acid hybrids (Azhikina et al., 1993; Sproat & Lamond, 1993). The A-T′ base pair exhibits the same hydrogen-bonding pattern as A-T, and its introduction into a hybrid usually increases the T_m by a small amount (Connolly & Newman, 1989; Newman et al., 1990; Kuimelis & Nambiar, 1994). By contrast, the A′-T′ doublet should act like a mismatch. Model building indicates that steric clash between the 2-thio group of thymine and the 2-amino group of adenine tilts the bases relative to each other, thereby allowing only one hydrogen bond to form.

Hybridization Properties of SBC ODNs. On the basis of the hydrogen-bonding properties of 2-aminoadenine and 2-thiothymine, we expected complementary ODNs which contained these bases to exhibit SBC properties. For the studies described here, two complementary 20-mer sequences (60% A/T) were synthesized with regular or modified adenine and thymine bases to generate the eight ODNs listed in Figure 3. Each of the 16 possible hybrids was formed in

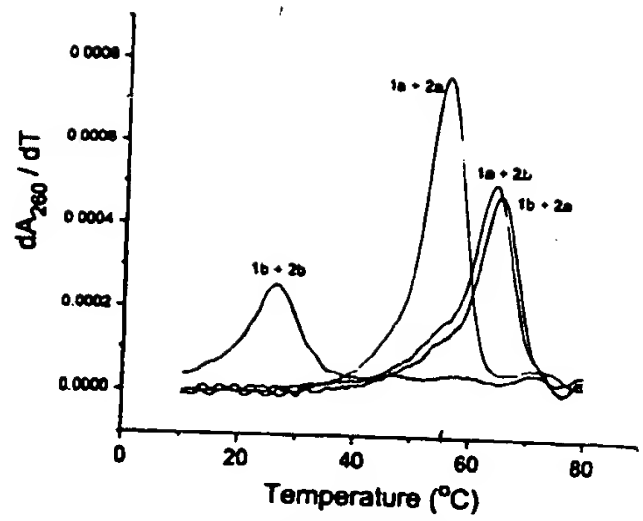


FIGURE 4: First-derivative melting curves of representative hybrids.

10 mM Na_2HPO_4 (pH 7.0) buffer containing 200 mM NaCl and 0.1 mM EDTA and assayed by ultraviolet monitored thermal denaturation.

Representative first-derivative melting curves are shown in Figure 4. The DNA–DNA hybrid (**1a**–**2a**) had a T_m of 55 °C, whereas the two alternative SBC–DNA hybrids (**1a**–**2b** and **1b**–**2a**) had T_m values approximately 10 °C higher. The relatively sharp, symmetric peaks obtained for the SBC–DNA hybrids indicated that A′-T and A-T′ base pairs did not alter the cooperativity of hybridization. The SBC–SBC hybrid (**1b**–**2b**) had a T_m of 26 °C, nearly 30 °C lower than that of the DNA–DNA hybrid and 40 °C lower than those of the two SBC–DNA hybrids. Above approximately 35 °C, the two complementary SBC 20-mers coexisted as single-stranded molecules capable of forming stable fully base-paired hybrids with normal complements. By these criteria, they met our definition of selectively binding complementary ODNs. The destabilizing effect of a A′-T′ doublet relative to an authentic mismatch was estimated in separate experiments which showed that the T_m of the control duplex (**1a**–**2a**) was on average depressed 2.4 °C for each introduction of a A′-T′ doublet versus 6.3 °C for each introduction of a T-T mismatch.

The free energy ($\Delta G°$) of each hybrid was determined using the two-state approximation for melting (Petershiem & Turner, 1983). These values are summarized together with the T_m results in Table 1. At 37 °C, the two SBC–DNA hybrids (**1a**–**2b** and **1b**–**2a**) were 1 kcal/mol more stable than the DNA–DNA hybrid (**1a**–**2a**) and 11 kcal/mol more stable than the SBC–SBC hybrid (**1b**–**2b**). On the basis of these differences, the equilibrium binding constant for the SBC–DNA duplexes was nearly 8 orders of magnitude greater than that for the SBC–SBC duplex. The free energy values for the other hybrids in Table 1 were as anticipated and further confirmed that the A′-T, A-T′, and A′-T′ base juxtapositions, respectively, stabilized, had little effect on, or destabilized the duplex under study.

Formation and Stability of Mobile Junctions. The hybridization properties of SBC ODNs **1b** and **2b** should favor the formation of stable, mobile three-way junctions between these ODNs and longer unmodified duplexes which contain the same sequences at one end. Two 40-mer hybrids (see Figure 3) were prepared to test this assumption, one with a blunt end (hybrid **3**–**4**) and the other with a five-base long recessed end (hybrid **3**–**5**). Three-arm junctions were formed two different ways. In the first protocol, each member strand of the longer duplex was separately hybrid-

Selectively

Table 1: Mel
Hybrids Subs
Base Analogs

hybrid	A
1a–2a	
1b–2a	
1a–2b	
1b–2b	
1a–2c	
1a–2d	
1b–2c	
1b–2d	
1c–2a	
1c–2b	
1c–2c	
1c–2d	
1d–2a	
1d–2b	
1d–2c	
1d–2d	

[a] Determine
Na_2HPO_4 (pH
presence of A
indicated. [c] Ea
of at least thre
in free energie

ized to the c
both partial
junction. S
mation of oth
ODNs. In t
preformed an
two invading
and **3**–**5** (re
ensure that a
ated by stra
unmodified (
molecule for

A gel mobi
junction form
Sequential hy
a good yield
same protoco
of the desirec
adenine for ac
formation of n
the yield (lan
SBC or norm
product (lanes
reactions app
migration to
stranded 40-m

The stabilit
whether norm
hybridization
modified ODI
resolve the fou
20-mer duple
and 4 of Figu
during the cou
arm junction f
did not underg
5). This was e

Table 1: Melting Transition Temperatures and Free Energies for Hybrids Substituted with 2-Aminoadenine and/or 2-Thiothymine Base Analogs[a]

hybrid	A-T doublets[b]	T_m[c] (°C)	$-\Delta G°_{37}$[c] (kcal/mol)	$-\Delta G°_{60}$[c] (kcal/mol)
1a−2a	A-T	55	16.2	8.1
1b−2a	A-T′, A′-T	65	17.3	11.2
1a−2b	A-T′, A′-T	63	17.2	10.7
1b−2b	A′-T′	26	6.2	0.3
1a−2c	A-T, A′-T	59	17.4	9.3
1a−2d	A-T, A-T′	58	16.6	9.4
1b−2c	A′-T, A′-T′	51	13.2	7.0
1b−2d	A-T′, A′-T′	41	10.7	4.0
1c−2a	A-T, A′-T	60	18.8	10.0
1c−2b	A′-T, A′-T′	42	10.8	4.2
1c−2c	A′-T	66	20.3	12.1
1c−2d	A-T, A′-T′	35	8.6	2.3
1d−2a	A-T, A-T′	57	13.6	9.5
1d−2b	A-T′, A′-T′	47	11.3	6.9
1d−2c	A-T, A′-T′	41	10.1	5.4
1d−2d	A-T′	61	14.1	10.0

[a] Determined in 200 mM NaCl, 0.1 mM EDTA, and 10 mM Na_2HPO_4 (pH 7.0) with a 16 μM total nucleotide concentration. [b] The presence of A-T, A′-T, A-T′, and A′-T′ doublets in each hybrid is indicated. [c] Each reported value for T_m and free energy is an average of at least three separate experiments; uncertainties in T_m values and in free energies are estimated at ±1.0 °C and ±15%, respectively.

ized to the complementary "invading" 20-mer, after which both partial hybrids were combined to form a three-arm junction. Sequential hybridization permitted the formation of otherwise unstable junctions involving unmodified ODNs. In the second protocol, the longer duplex was preformed and then incubated at room temperature with the two invading 20-mers. The high T_m values of hybrids **3**−**4** and **3**−**5** (respectively, 69 and 68 °C) were expected to ensure that any three-arm junction would have been generated by strand invasion. Both SBC (**1b** and **2b**) and unmodified (**1a** and **2a**) ODNs were tested for branched molecule formation.

A gel mobility shift analysis was used to detect three-arm junction formation with the blunt-ended hybrid (Figure 5). Sequential hybridization using the SBC 20-mers generated a good yield of the three-arm junction (lane 5), whereas the same protocol using normal 20-mers generated much less of the desired junction (lane 3). Substitution of 2-aminoadenine for adenine in the unmodified ODNs to promote the formation of more stable branched molecules did not improve the yield (lane 4). Sequential hybridization using only one SBC or normal 20-mer failed to generate any branched product (lanes 7 and 8). The joint molecule formed in these reactions appears to have undergone spontaneous branch migration to yield a single-stranded 20-mer and a double-stranded 40-mer (Radding et al., 1977).

The stability of the three-arm junction depended upon whether normal or SBC 20-mers were used in the sequential hybridization protocol. When the junction contained unmodified ODNs, branch migration required many hours to resolve the four component strands into separate 40-mer and 20-mer duplexes (data not shown). The smear in lanes 3 and 4 of Figure 5 is attributable to this reaction occurring during the course of electrophoresis. By contrast, the three-arm junction formed using SBC ODNs was very stable and did not undergo branch migration (lanes 5 and 6 in Figure 5). This was expected since resolution of the junction would have been accompanied by a reduction in the total number of base pairs.

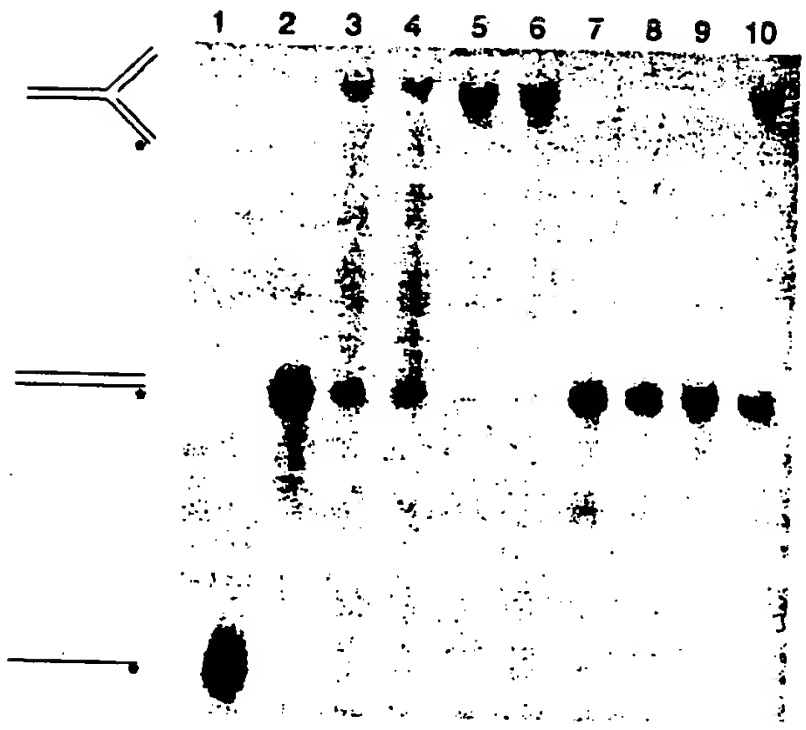


FIGURE 5: Gel shift analysis of branched molecules formed by the interaction of normal or SBC 20-mers with the blunt-ended duplex **3**−**4**. Reactions were carried out at room temperature (25−27 °C) in the standard buffer. ODN **3** was end-labeled and present at a final concentration of 5×10^{-7} M. When used in forming a branched molecule in this or other experiments, the molar ratio of ODNs **3**, **4**, **1a**−**c**, and **2a**−**c** was 1:1:5:5. Intermediate hybrids used in the sequential hybridization protocols were formed by incubating the respective strands for at least 10 min. Controls: lane 1, ODN **3**; lane 2, duplex **3**−**4**. Sequential hybridization of ODNs to form a three-arm junction followed by a 15 min incubation: lane 3, duplex **3**−**2a** added to duplex **4**−**1a**; lane 4, duplex **3**−**2c** added to duplex **4**−**1c**; lane 5, duplex **3**−**2b** added to duplex **4**−**1b**. Sequential hybridization of ODNs to form a three-arm junction followed by a 12 h incubation: lane 6, duplex **3**−**2b** added to duplex **4**−**1b**. Sequential hybridization to form a D-loop-like molecule: lane 7, ODN **4** added to duplex **3**−**2a**; lane 8, ODN **4** added to duplex **3**−**2b**. Strand invasion of duplex **3**−**4** to form a three-arm junction (12 h incubation): lane 9, ODNs **1a** and **2a**; lane 10, ODNs **1b** and **2b**. At the completion of each reaction, an aliquot was quickly frozen in dry ice and stored at −20 °C until it was analyzed by electrophoresis in a precooled (10 °C) 8% nondenaturing polyacrylamide gel run at room temperature. The closely spaced doublets in lanes 3 and 4 probably represent different conformers of the same three-arm junction; such doublets have been reported previously (Zhong et al., 1994).

A thermostability experiment showed that 50% of the three-arm junction formed between the SBC 20-mers and the blunt-ended 40-mer hybrid had resolved into the component duplex and single strands at 62 °C (Figure 6). This apparent T_m was only somewhat lower than the T_m of the blunt-ended 40-mer duplex in the same buffer and explained why the amount of three-arm junction increased above 65 °C. At those temperatures, both the three-arm junction and the underlying 40-mer duplex underwent denaturation. When aliquots of these solutions were rapidly cooled for storage, the three-arm junctions reformed.

Strand invasion of preformed hybrids was only observed with the SBC ODNs. These complementary ODNs formed stable three-arm junctions with both the blunt-ended hybrid (Figure 5, lane 10) and the recessed hybrid (Figure 7, left lanes 1−5). By providing an annealing site for one of the SBC ODNs, the recessed hybrid was more rapidly invaded. When incubated with the same hybrids, normal ODNs were devoid of strand invasion activity (Figure 5, lane 9; Figure

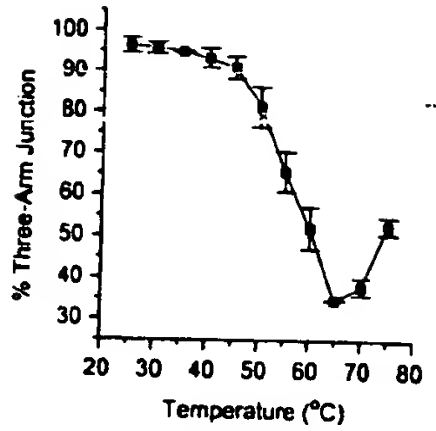


FIGURE 6: Thermostability of the three-arm junction formed between the blunt-ended duplex **3**−**4** and paired SBC 20-mers. The three-arm junction was formed by sequential hybridization of SBC ODNs **1b** and **2b** with duplex strands **3** and **4** as described in Figure 5. The temperature of the solution containing the junction was then raised in 5 °C increments starting from 25 °C. After 10 min of incubation at each temperature, an aliquot was quickly frozen in dry ice and stored at −20 °C while awaiting gel analysis. Bands corresponding to the three-arm junction and 40-mer duplex were detectable. The percent of three-arm junction was determined using a phorphorimager. Identical results were obtained when reaction aliquots were immediately run on a gel without prior storage.

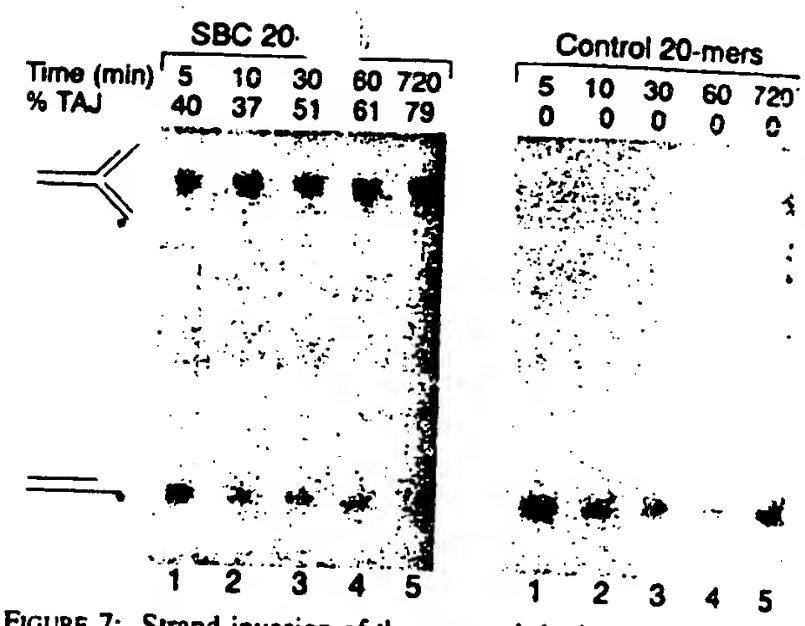


FIGURE 7: Strand invasion of the recessed duplex **3**−**5** by paired SBC or normal 20-mers. Reaction conditions and electrophoretic analysis were as described in Figure 5. The **3**−**5** duplex (4 × 10⁻⁸ M) was incubated with a 5-fold molar excess of normal ODNs **1a** and **2a** or SBC ODNs **1b** and **2b**. Aliquots were removed at the indicated times to determine the extent of strand invasion. The percent yield of three-arm junction was determined by phosphorimage analysis of the dried gel: left lanes 1−5, SBC ODNs; right lanes 1−5, normal ODNs. TAJ = three-arm junction.

7, right lanes 1−5). This was not surprising since these complementary ODNs hybridized to each other.

DISCUSSION

The single-stranded character of paired SBC ODNs is key to their remarkable strand invasion properties. At the concentrations used here, paired SBC ODNs can readily anneal to a homologous duplex which contains a single-stranded overhang or a transiently frayed blunt end. Invasion of the duplex by both ODNs follows. Use of a single self-complementary SBC ODN would reduce the molecularity of the annealing step and further promote strand invasion. In either case, strand invasion is energetically favored by the creation of one additional base pair for each step of invasion and may be kinetically favored by the unorthodox structure of the branch migration intermediate. Normally, strand exchange between two duplexes occurs within the context of a Holliday junction. In the presence of $MgCl_2$, the four double-stranded arms of a Holliday junction form two quasicontinuous helices (Lilley & Clegg, 1993) which retard branch migration (Panyutin & Hsieh, 1994; Panyutin et al., 1995). By contrast, the unusual junction formed when using SBC ODNs contains three double-stranded and two single-stranded arms (see the intermediate in Figure 1B). We hypothesize that the absence of a fourth duplex arm will increase conformational freedom at the crossover point and promote strand invasion both in the presence and in the absence of $MgCl_2$. The rate and extent of strand invasion observed here would have been improved by conducting the experiments at 37 °C instead of at room temperature. At the higher temperature, the paired SBC ODNs would have been totally single-stranded and the ends of the duplex more frayed.

Using two different protocols, the paired SBC ODNs formed stable, mobile three-arm junctions whereas the paired unmodified ODNs did not. While the SBC-containing junctions were slightly more stable than the unmodified junctions due to the presence of A′-T base pairs, we believe that the susceptibility of three-arm junctions to resolution is primarily determined by the relative base pairing count of branched versus linear associations of the strands. Extrusion of two SBC ODNs from a three-arm junction would be accompanied by a reduction in the total number of base pairs, making the branched molecule the more stable species despite the presence of a junction. On the other hand, extrusion of two normal ODNs from a three-arm junction should occur with no loss of base pairing. In this more typical situation, the three-arm junction would be less stable than the two component linear duplexes (Lu et al., 1991; Leontis et al., 1994) and susceptible to resolution (Panyutin & Hsieh, 1993).

For optimal strand invasion of a short duplex, each of the two SBC ODNs should form a hybrid with the respective complementary strand which is fully base-paired and has equal or greater stability than the starting duplex. Clearly, SBC ODNs with A′ and T′ substitutions meet these criteria. By contrast, two complementary ODNs with a limited number of mismatched bases at mutually exclusive positions might not interact with each other and yet form weak hybrids with complements possessing the wild-type sequences. Such ODNs would not likely strand invade and do not meet the definition of SBC ODNs. The robustness of a pair of SBC ODNs can be quantified by comparing the difference in stability between the normal hybrid and the SBC−SBC hybrid to the sum of the stability changes observed when only the Watson or the Crick strand of the hybrid is substituted with an SBC ODN. The greater this difference the more powerful the SBC properties.

Rules for the design of SBC ODNs using A′ and T′ bases have yet to be defined. On the basis of the free energy values in Table 1, one can assume that each A′-T′ doublet in an SBC−SBC duplex contributes 0.8 kcal/mol of destabilization. This assumption provides a framework for predicting whether a given pair of ODNs can exhibit SBC properties. Both length and sequence will be important factors in determining the T_m of paired oligomeric complements. It is uncertain whether 2-aminoadenine will compromise the specificity of SBC ODNs. Although capable of degenerate

base pairing (Cheong et al., 1988), this analog is used in place of adenine by the S-2L cyanophage (Kirnos et al., 1977; Khudyakov et al., 1978), and the triphosphate of 2-amino-adenine acts as a true analog of ATP in transcription (Rackwitz & Scheit, 1977).

The design of the A′-T′ base pair has taken advantage of the absence of a functional group at the 2 position of adenine. Introduction of an amino group at this position together with substitution of sulfur for oxygen at the same position of thymine has provided two base analogs which permit substituted ODNs to discriminate between modified and natural complements. The design of G/C analogs for use in SBC ODNs must of necessity use different strategies. The synthesis and properties of ODNs substituted with such analogs will be described elsewhere (Woo et al., 1996).

The model strand invasion substrates employed in this study can be likened to the stem of a hairpin which might be found in a long single-stranded nucleic acid. While a traditional ODN is usually ill equipped to deal with such a structure, we have shown that paired SBC ODNs can readily invade a short duplex. In theory, paired SBC ODNs should not pay a penalty for binding to structured regions in single-stranded DNA or RNA due to a net increase in base pairing. By contrast, the binding of a traditional ODN to a long target is inhibited by the presence of secondary structure. From this analysis, it follows that the relative advantage of using paired SBC ODNs over a traditional ODN in hybridizing to a long single-stranded nucleic acid will be determined by the degree of secondary structure at the hybridization site, with localized hairpins providing the greatest advantage.

The inherent stability of RNA duplexes and the ability of RNA to form G-U base pairs favor intramolecular base pairing of these molecules within the cell. As a result, transcripts exist as highly folded, globular structures in which a large fraction of the bases are hydrogen-bonded (Laptev et al., 1993; Uhlenbeck, 1995). The preponderance of double-stranded segments in mRNA may explain why certain sense ODNs can act as antisense agents (Colige et al., 1993; Thierry et al., 1993). Generally, RNA folding programs cannot accurately predict the native structure of such long molecules because each can adopt many alternative conformations with nearly equivalent stabilities. This has hindered the development of antisense technology by making the selection of single-stranded regions in mRNA a largely empirical process.

The results presented here suggest that paired complementary SBC ODNs could be used as effective antisense agents. In this regard, paired SBC ODNs should be more tolerant of secondary structure features than a regular antisense ODN. Furthermore, SBC−RNA hybrids should be stabilized by the known propensity of A′- or T′-containing ODNs to form A-type duplexes (Connolly & Newman, 1989; Newman et al., 1990; Garriga et al., 1993). We have initiated studies using SBC ODNs to strand invade naturally occurring hairpin structures in RNA. Synthesis of SBC ODNs with modified RNA backbones should be chemically straightforward and may potentiate their use as antisense agents.

Besides hybridizing to single-stranded nucleic acids, paired SBC ODNs might also be used to target any sequence in a long double-stranded DNA. In this respect, strand invasion of SBC ODNs into the end of a duplex bears some similarity to the branch capture reaction (Quartin et al., 1989; Weinstock & Wetmur, 1990). The more general case of targeting an internal sequence distant from an end would be feasible if paired SBC ODNs can be used by the recA protein. Others have shown that this recombinase can catalyze the hybridization of both strands of a denatured restriction fragment to the complementary sequences in a duplex substrate (Jayasena & Johnston, 1993; Sena & Zarling, 1993). The resultant double D-loop is a stable structure which survives the removal of recA. While this technique does not work with short oligomeric complements, it might utilize paired SBC ODNs.

ACKNOWLEDGMENT

We thank Drs. S. Freier, P. Hsieh, and N. Leontis for critically reading the manuscript. The early work on SBC ODNs by our colleague Dr. J. Woo served as the stimulus for the study reported here.

REFERENCES

Afonina, I., Kutyavin, I., Lukhtanov, E., Meyer, R. B., & Gamper, H. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 3199−3204.

Asseline, U., Delarue, M., Lancelot, G., Toulme, F., Thuong, N. T., Montenay-Garestier, T., & Helene, C. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3297−3301.

Azhikina, T., Veselovskaya, S., Myasnikov, V., Potapov, V., Ermolayeva, O., & Sverdlov, E. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 11460−11462.

Brossalina, E., & Toulme, J.-J. (1993) *J. Am. Chem. Soc. 115*, 796−797.

Chastain, M., & Tinoco, I., Jr. (1993) in *Antisense Research and Applications* (Crooke, S. T., & Lebleu, B., Eds.) pp 55−66, CRC Press, Boca Raton, FL.

Cheong, C., Tinoco, I., Jr., & Chollet, A. (1988) *Nucleic Acids Res. 16*, 5115−5122.

Colige, A., Sokolov, B. P., Nugent, P., Baserga, R., & Prockop, D. J. (1993) *Biochemistry 32*, 7−11.

Connolly, B. A., & Newman, P. C. (1989) *Nucleic Acids Res. 17*, 4957−4974.

Ecker, D. J., Vickers, T. A., Bruice, T. W., Freier, S. M., Jenison, R. D., Manoharan, M., & Zounes, M. (1992) *Science 257*, 958−961.

Fathi, R., Goswami, B., Kung, P.-P., Gaffney, B. L., & Jones, R. A. (1990) *Tetrahedron Lett. 31*, 319−322.

Francois, J.-C., & Helene, C. (1995) *Biochemistry 34*, 65−72.

Francois, J.-C., Thuong, N. T., & Helene, C. (1994) *Nucleic Acids Res. 22*, 3943−3950.

Gamper, H. B., Cimino, G. D., & Hearst, J. E. (1987) *J. Mol. Biol. 197*, 349−362.

Gamper, H. B., Reed, M. W., Cox, T., Virosco, J. S., Adams, A. D., Gall, A. A., Scholler, J. K., & Meyer, R. B., Jr. (1993) *Nucleic Acids Res. 21*, 145−150.

Garriga, P., Garcia-Quintana, D., Sagi, J., & Manyosa, J. (1993) *Biochemistry 32*, 1067−1071.

Gryaznov, S., & Chen, J.-K. (1994) *J. Am. Chem. Soc. 116*, 3143−3144.

Howard, F. B., & Miles, H. T. (1984) *Biochemistry 23*, 6723−6732.

Jayasena, V. K., & Johnston, B. H. (1993) *J. Mol. Biol. 230*, 1015−1024.

Khudyakov, I. Y., Kirnos, M. D., Alexandrushkina, N. I., & Vanyushin, B. F. (1978) *Virology 88*, 8−18.

Kirnos, M. D., Khudyakov, I. Y., Alexandrushkina, N. I., & Vanyushin, B. F. (1977) *Nature 270*, 369−370.

Kuimelis, R. G., & Nambiar, K. P. (1994) *Nucleic Acids Res. 22*, 1429−1436.

Kutyavin, I. V., Podyminogin, M. A., Bazhina, Y. N., Fedorova, O. S., Knorre, D. G., Levina, A. S., Mamayev, S. V., & Zarytova, V. F. (1988) *FEBS Lett. 238*, 35−38.

Lamm, G. M., Blencowe, B. J., Sproat, B. S., Iribarren, A. M., Ryder, U., & Lamond, A. I. (1991) *Nucleic Acids Res. 19*, 3193−3198.

Laptev, A. V., Lu, Z., Colige, A., & Prockop, D. J. (1994) *Biochemistry 33*, 11033−11039.
Leontis, N. B., Hills, M. T., Piotto, M., Ouporov, I. V., Malhotra, A., & Gorenstein, D. G. (1994) *Biophys. J. 68*, 251−265.
Lilley, D. M. J., & Clegg, R. M. (1993) *Annu. Rev. Biophys. Biomol. Struct. 22*, 299−328.
Lima, W. F., Monia, B. P., Ecker, D. J., & Freier, S. M. (1992) *Biochemistry 31*, 12055−12061.
Lu, M., Guo, Q., & Kallenbach, N. R. (1991) *Biochemistry 30*, 5815−5820.
Monia, B. P., Lesnik, E. A., Gonzalez, C., Lima, W. F., McGee, D., Guinosso, C. J., Kawasaki, A. M., Cook, P. D., & Freier, S. M. (1993) *J. Biol. Chem. 268*, 14514−14522.
Newman, P. C., Nwosu, V. U., Williams, D. M., Cosstick, R., Seela, F., & Connolly, B. A. (1990) *Biochemistry 29*, 9891−9901.
Nielsen, P. E., Egholm, M., & Buchardt, O. (1994) *Bioconjugate Chem. 5*, 3−7.
Panyutin, I. G., & Hsieh, P. (1993) *J. Mol. Biol. 230*, 413−424.
Panyutin, I. G., & Hsieh, P. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2021−2025.
Panyutin, I. G., Biswas, I., & Hsieh, P. (1995) *EMBO J. 14*, 1819−1826.
Petershiem, M., & Turner, D. H. (1983) *Biochemistry 22*, 256−263.
Polushin, N. N., & Cohen, J. S. (1994) *Nucleic Acids Res. 22*, 5492−5496.
Prakash, G., & Kool, E. T. (1991) *J. Chem. Soc., Chem. Commun.*, 1161−1162.
Quartin, R. S., Plewinska, M., & Wetmur, J. G. (1989) *Biochemistry 28*, 8676−8682.
Rackwitz, H. R., & Scheit, K. H. (1977) *Eur. J. Biochem. 72*, 191−200.
Radding, C. M., Beattie, K. L., Holloman, W. K., & Wiegand, R. C. (1977) *J. Mol. Biol. 116*, 825−839.
Richardson, P. L., & Schepartz, A. (1991) *J. Am. Chem. Soc. 113*, 5109−5111.
Sena, E. P., & Zarling, D. A. (1993) *Nat. Genet. 3*, 365−372.
Sproat, B. S., & Lamond, A. I. (1993) in *Antisense Research and Applications* (Crooke, S. T., & Lebleu, B., Eds.) pp 352−362. CRC Press, Boca Raton, FL.
Sproat, B. S., Iribarren, A. M., Garcia, R. G., & Beijer, B. (1991) *Nucleic Acids Res. 19*, 733−738.
Thierry, A. R., Rahman, A., & Dritschilo, A. (1993) *Biochem. Biophys. Res. Commun. 190*, 952−960.
Uhlenbeck, O. C. (1995) *RNA 1*, 4−6.
Wagner, R. W., Matteucci, M. D., Lewis, J. G., Gutierrez, A. J., Moulds, C., & Froehler, B. C. (1993) *Science 260*, 1510−1513.
Wang, S., & Kool, E. T. (1994) *J. Am. Chem. Soc. 116*, 8857−8858.
Weinstock, P. H., & Wetmur, J. G. (1990) *Nucleic Acids Res. 18*, 4207−4213.
Woo, J., Meyer, R. B., & Gamper, H. B. (1996) *Nucleic Acids Res. 24*, 2470−2475.
Zhong, M., Rashes, M. S., Leontis, N. B., & Kallenbach, N. R (1994) *Biochemistry 33*, 3660−3667.

BI960626V

Influ

Depart

ABSTRA
This an
covalen
topoisor
either ϕ
was use
Moreov
the topo
topoisor
trapping
on DNA
DNA, w

Elsamicin A (
nds specificall
995; Párraga &
d is capable c
e B-form (Jime
causes DNA b
reducing agent
tivity appears
peroxide dism
at the aglycon
dicals. Elsamic
veral murine
dergone phase
mitrucin (Goss
count for the a
nsidering its pc
merases, inclu
portant for ma
mal DNA duri
plication (Chen
terization of th
s shown that it
be a more pot
) or adriamycin
y be the most
scribed (Lorico
whether this e
ect on the topo
e elsamicin ma
NA complexes, t
dicases (Bachur

This work was fi
n, and the suppor
ugh the 'Centre d
To whom corresp
Biología Molecula
C, J. Girona-Salg
ne: 34-3-400 61
sic.es.
Abstract publishe

2470–2475 Nucleic Acids Research, 1996, Vol. 24, No. 13

G/C-modified oligodeoxynucleotides with selective complementarity: synthesis and hybridization properties

Jinsuk Woo, Rich B. Meyer, Jr and Howard B. Gamper*

Epoch Pharmaceuticals, Inc., 1725 220th Street SE, Bothell, WA 98021, USA

Received April 29, 1996; Accepted May 22, 1996

ABSTRACT

Modified oligodeoxyribonucleotides (ODNs) that have unique hybridization properties were designed and synthesized for the first time. These ODNs, called selective binding complementary ODNs (SBC ODNs), are unable to form stable hybrids with each other, yet are able to form stable, sequence specific hybrids with complementary unmodified strands of nucleic acid. To make SBC ODNs, deoxyguanosine (dG) and deoxycytidine (dC) were substituted with deoxyinosine (dI) and 3-(2′-deoxy-β-D-ribofuranosyl)pyrrolo-[2,3-*d*]-pyrimidine-2-(3*H*)-one (dP), respectively. The hybridization properties of several otherwise identical complementary ODNs containing one or both of these nucleoside analogs were studied by both UV monitored thermal denaturation and non-denaturing PAGE. The data showed that while dI and dP did form base pairs with dC and dG, respectively, dI did not form a stable base pair with dP. A self-complementary ODN uniformly substituted with dI and dP acquired single-stranded character and was able to strand invade the end of a duplex DNA better than an unsubstituted ODN. This observation implies that SBC ODNs should effectively hybridize to hairpins present in single-stranded DNA or RNA.

INTRODUCTION

Oligodeoxyribonucleotides (ODNs) do not effectively hybridize to complementary sequences which are already base paired. Without the assistance of recombinase enzymes such as recA (1), accessibility of ODNs to double-stranded DNA (dsDNA) is usually restricted to homopurine runs (2) or to extruded single-stranded sequences in supercoiled DNA (3). Although less of an issue with single-stranded DNA (ssDNA) or RNA, hybridization of ODNs to many sequences in these molecules can be compromised by intramolecular base pairing (4,5). While numerous hybridization strategies have been described to overcome or exploit secondary structure, none provides a general solution to the problem. Examples include modified ODNs which form unusually stable hybrids (6–12), ODNs which form triple-stranded complexes (13), ODNs which hybridize to hairpins or contiguous flanking sequences (14–18), and the use of 'effector' ODNs (19) and 'tethered' ODNs (20) to improve binding affinity through cooperative interactions.

A pair of uniquely modified complementary ODNs (or a single self-complementary ODN) that do not hybridize to each other, yet do hybridize to unmodified complementary sequences might offer a general solution to the challenge of targeting any site in DNA or RNA. If such a pair of ODNs could be synapsed to a homologous region in dsDNA by recombination, a complement-stabilized or double D-loop (21–22) would be formed (Fig. 1a). Unlike a simple D-loop, the double D-loop is relatively stable and might inhibit gene expression. Alternatively, the same type of paired ODNs could be hybridized to a unique sequence in long, single-stranded nucleic acid. To the extent that sequence is involved in secondary structure (such as a localized hairpin; Fig. 1b), the paired ODNs should have an advantage over a standard ODN. Whether such ODNs are used as probes or antisense agents, their hybridization to a target should generate more new base pairs than an unmodified ODN. This is depicted in Figure 1.

We describe the synthesis of a cytosine (dC) analog. When incorporated into an ODN it rearranged to a nucleoside (**dP**) which formed 2-3 hydrogen bonds when opposite a guanosine (dG) and 1-2 hydrogen bonds when opposite an inosine (**dI**). When every dC and dG in a pair of complementary ODNs was substituted with **dP** and **dI**, respectively, the ODNs did not hybridize to each other yet did hybridize to unmodified, complementary ODNs. By this criterion, the ODNs demonstrated selective binding complementarity and are designated SBC ODNs. Although the SBC–DNA hybrids were less stable than the DNA–DNA hybrid, a self-complementary SBC ODN was more effective than the corresponding unmodified ODN in strand invading a homologous duplex DNA. Further development of the SBC concept will depend upon the synthesis of base analogs which form stronger pairs with the natural complement.

MATERIALS AND METHODS

Materials and their sources were as follows: DNA synthesis reagents, Glen Research; phosphodiesterase I (*Crotalus adamanteus* venom), alkaline phosphatase (calf intestinal) and DNase I, Amersham Life Science; T4 polynucleotide kinase (10 U/μl),

* To whom correspondence should be addressed

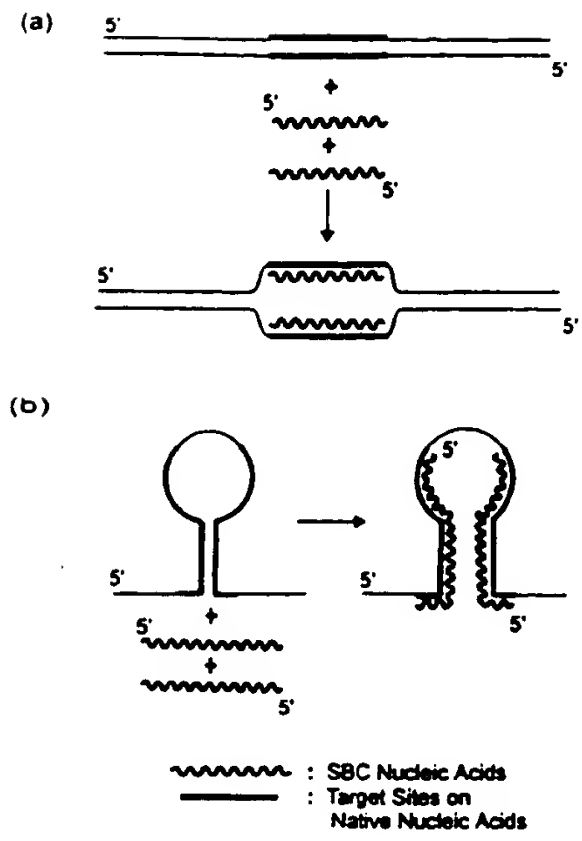


Figure 1. The possible applications of SBC ODNs. (a) The interaction of SBC ODNs with dsDNA to form a complement stabilized D-loop in the presence of a recombinase such as recA. (b) The strand invasion of a DNA or RNA hairpin by SBC ODNs. In each example hybridization leads to an increase in the total number of base pairs, thus providing a thermodynamic drive for the reaction.

Promega: [γ-^{32}P]ATP, NEN Research. Commercial reagents were used as received. ^{1}H-NMR spectra were determined on a Varian Gemini-300. Elemental analysis was performed by Quantitative Technologies Inc. (Whitehouse, NJ). UV spectra were measured on a Beckman DU-40 spectrophotometer or a Perkin Elmer Lamda 2S UV/VIS spectrophotometer.

Preparation of 3-(2′-deoxy-β-D-ribofuranosyl)furano-[2,3-*d*]-pyrimidine-6(5*H*)-one (dF)

5–Ethynyl–2′–deoxyuridine (3 g, 11.9 mmol) (23) and copper (I) iodide (500 mg, 2.6 mmol) in a 250 ml two–necked round–bottomed flask were dried in vacuo for 3 h, placed under argon, and suspended in anhydrous DMF (35 ml) and triethylamine (15 ml). The solution was vigorously stirred at 120°C under argon and every 30 min fresh copper (I) iodide (250 mg, 1.3 mmol) was added until most of the starting material had reacted. After 2 h, the resulting mixture was filtered and the filtrate was concentrated *in vacuo* to dryness. The residue was suspended in acetone (100 ml) and stirred overnight. The desired product was filtered, washed with acetone (20 ml), and dried *in vacuo* to afford 2.2 g of **dF** as a slightly yellowish solid. The remaining product in mother liquor was further purified by silica gel column chromatography (elution solvent: 25% MeOH in EtOAc) to afford an additional 0.3 g of **dF** (total yield: 2.5 g, 83%): mp 167–168°C; UV (0.05 M $KHPO_4$/NaOH, pH 7) λ_{max} 322 nm (ε 12 500). Anal. calcd for $C_{11}H_{12}N_2O_5$: C, 52.38; H, 4.80; N, 11.11. Found: C, 52.11; H, 4.81; N, 10.91. ^{1}H NMR (DMSO-d_6): the same as reported by Kumar *et al.* (24).

Preparation of 5′-*O*-(4,4′-dimethoxytrityl)-dF

dF (2.17 g, 8.6 mmol) was dried *in vacuo* at 60°C overnight and then added to 4,4′-dimethoxytrityl-chloride (3.51 g, 10.4 mmol) and anhydrous triethylamine (2.4 ml) in pyridine (30 ml). After 2 h at room temperature under argon, the resulting mixture was diluted with an equal volume of water and extracted with two 150 ml portions of ether. The ether layer was dried over anhydrous sodium sulfate and evaporated to dryness. The residue was dissolved in dichloromethane (20 ml) and the desired product (4.6 g) was precipitated by adding the solution to 400 ml of rapidly stirred hexanes. Filtration yielded 4.6 g (96%) of a white solid.

Preparation of 5′-*O*-(4,4′-dimethoxytrityl)-dF-3′-*O*-phosphoramidite

Chloro-[(β-cyanoethoxy)-*N*,*N*-diisopropylamino]-phosphine (2.9 g, 12.5 mmol) was added dropwise over 30 s to an anhydrous mixture of 5′-*O*-(4,4′-dimethoxytrityl)-**dF** (4.6 g, 8.3 mmol), diisopropylethyl amine (5.8 ml), and dichloromethane (27 ml) under argon (25). After 30 min at room temperature the reaction was stopped by adding anhydrous methanol (0.3 ml). The reaction mixture was extracted with 5% aqueous $NaHCO_3$ (2× 15 ml) and saturated aqueous NaCl (2× 15 ml). The organic layer was dried over anhydrous sodium sulfate, filtered and then evaporated under reduced pressure to afford a brown oil. This crude product was further purified by silica gel column chromatography using hexanes:CH_2Cl_2:EtOAc:Et_3N (4:3:2:1 by vol) as the solvent system. Fractions containing the desired product were combined, evaporated to dryness, and redissolved in EtOAc (10 ml). Precipitation from rapidly stirred hexanes (400 ml) yielded 5.9 g (94%) of purified material. ^{1}H NMR ($CDCl_3$) δ 8.88 (d, J = 18.6 Hz, 1H), 7.5–7.2 (m, 10H), 6.81 (m, 4H), 6.32 (m, 1H), 5.62 (d of d, J = 9.5, 3.9 Hz, 1H), 4.70 (m, 1H), 4.19 (m, 1H), 3.8–3.4 (m, 12H), 2.77 (m, 1H), 2.59 (t, J = 9.4 Hz, 2H), 2.44 (m, 2H), 1.25–1.0 (m, 12H).

Conversion of dF to dP

dF (1 g, 3.96 mmol) was dissolved in 30% aqueous ammonium hydroxide (30 ml). After overnight at room temperature, the resulting solution was concentrated *in vacuo* to dryness. The residue was suspended in acetone (50 ml), stirred overnight, and the undissolved product filtered to afford 850 mg. The mother liquor was concentrated to dryness and the residue was suspended in acetone (10 ml) overnight with stirring to yield an additional 100 mg of insoluble product (total 950 mg, 95.4%). This compound was analyzed by HPLC, UV and NMR and shown to be identical to authentic 3-(2′-deoxy-β-D-ribofuranosyl)pyrrolo-[2,3-*d*]-pyrimidine-2(3*H*)-one (**dP**).

Synthesis and purification of ODNs

ODNs containing modified bases were synthesized on 1 μmol scale using standard procedures for an ABI-394 DNA synthesizer. ODNs with the dimethoxytrityl group were purified by HPLC using a Hamilton PRP-1 (7.0 × 305 mm) reverse phase column employing a gradient of 5 to 45% CH_3CN in 0.1 M $Et_3NH^+OAc^-$, pH 7.5, over 20 min with a 2 ml/min flow rate. After detritylation with 80% acetic acid, the ODNs were precipitated by addition of 3 M sodium acetate and 1-butanol. The resulting ODNs were dried and further purified by using 20% denaturing PAGE as described by Hopkins *et al.* (26).

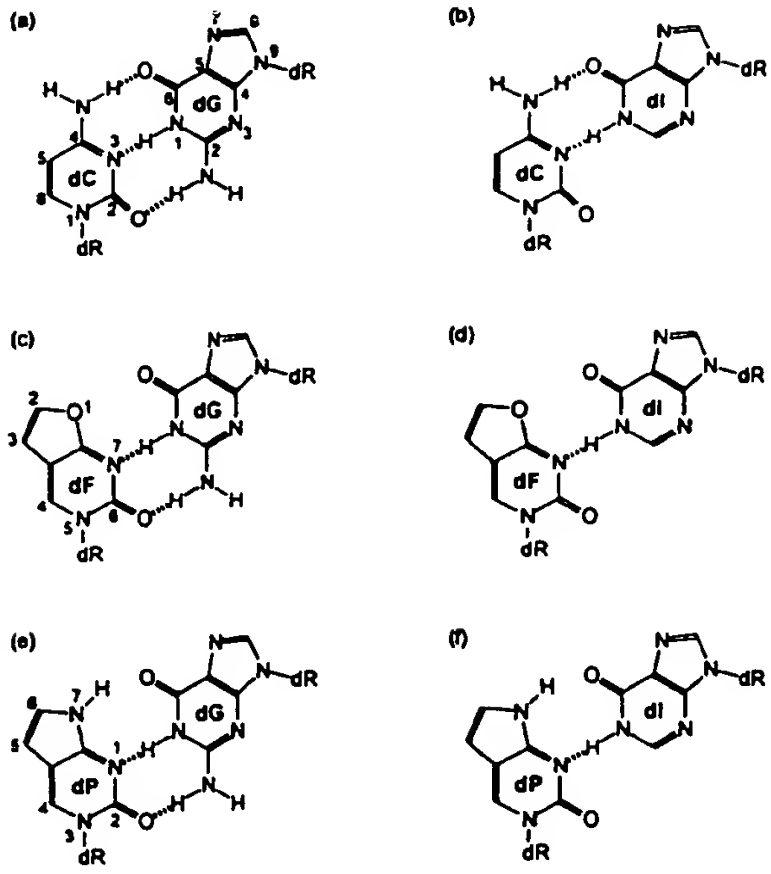


Figure 2. Base pairing schemes for dC and dG analogs.

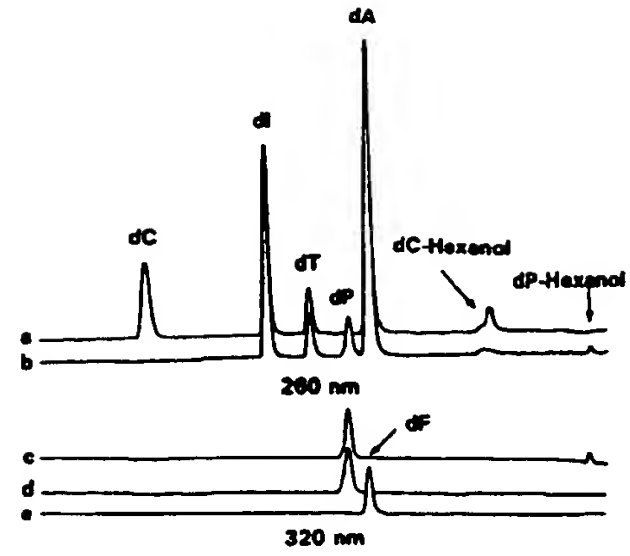


Figure 3. Reverse phase HPLC analysis: (a) enzymatic hydrolysate of Watson strand in **II** (see Table 1); (b and c) enzymatic hydrolysate of Watson strand in **VIII** (see Table 1); (d) authentic **dP**; (e) authentic **dF**. Detection was at 260 nm (a and b) or 320 nm (c, d and e). Retention times increase to the right.

Enzymatic digestion of ODNs

Enzymatic hydrolysis of ODNs was carried out as described by Woo *et al.* (27). The resulting hydrolysate was analyzed by HPLC with dual detection at 260 nm and 320 nm (Waters 994 Programmable Photodiode Array Detector) using a C-18 reverse phase column (Rainin, Microsorb™Short-One®). The solvent gradient was run at 1 ml/min as follows: solvent A, 0.1 M $Et_3NH^+OAc^-$, pH 7.5; solvent B, CH_3CN; a linear gradient 0 to 13% B over 10 min, a linear gradient to 100% B over 2 min, then isocratic 100% B for 3 min. Peaks were identified by comparison of retention times to those of authentic, commercial samples (dA, dG, dT and dC) and synthetic samples (**dF** and **dP**) prepared by known procedures (28).

Thermal denaturation data (T_m)

T_m values were recorded on a Perkin Elmer Lamda 2S UV/VIS spectrophotometer equipped with a temperature programmer (PTP-6) and interfaced to an IBM personal computer (PECSS software, Perkin Elmer). Scan rates were 0.5°C/min. Data were collected at 260 nm in the temperature range from 5 to 90°C. The T_m is defined as the temperature at half the maximal hyperchromicity using baseline correction at high and low temperature extremes (29). Samples were prepared by dissolving ODNs in TNM buffer [10 mM Tris–HCl (pH 8.0), 0.1 mM EDTA, 50 mM NaCl, 10 mM $MgCl_2$]. To ensure complete hybridization of complementary strands (1:1 molar ratio) before collecting data, the samples were incubated at 90°C for 2 min and cooled to 3°C over 1 h. The concentration of hybridized ODNs was approximately 2 μM.

Gel migration assay

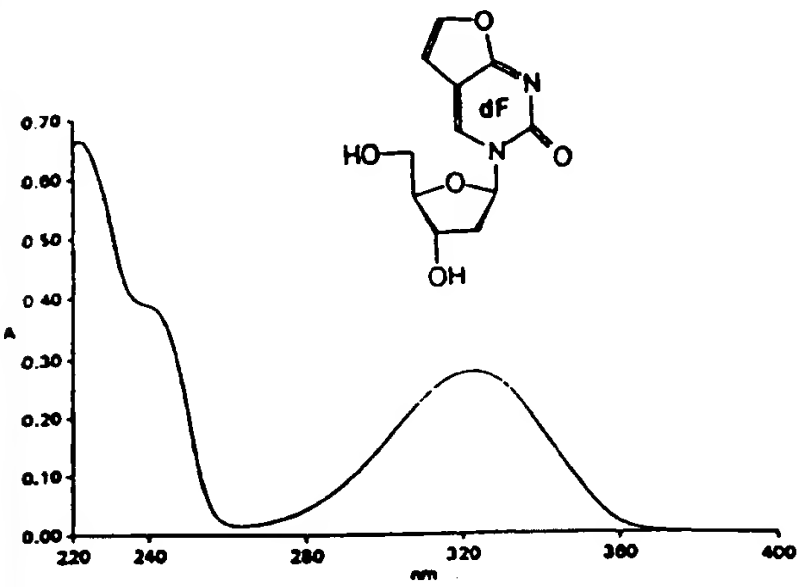
ODNs with an asterisk (*) in Figures 5 and 6 were 5′ ^{32}P-labeled using T4 kinase and [γ-^{32}P]ATP (30) and present at 0.5 μM unless otherwise indicated. Hybrids were formed by incubating the labeled ODN with a 2-fold molar excess of cold complementary ODN for 60 min at room temperature in 20 μl TNM buffer. These samples were then mixed with 20 μl loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, 2.5% Ficoll type 400) and then kept on ice prior to gel electrophoresis. Aliquots (5 μl) were analyzed in a 12% non-denaturing polyacrylamide gel [19:1 acrylamide: bisacrylamide, 0.35 mm thick, 20 × 16 cm, polymerized and run in TBE buffer (89 mM Tris-borate/2 mM EDTA) containing 3 mM $MgCl_2$]. Pre-electrophoresis in a BioRad Protean®II xi apparatus was performed for 1 h at 200 V and 10°C. Samples were loaded, and the gel was run as before until the bromophenol blue dye had traveled ~15 cm (~5 h). The gel was dried and visualized with a phosphorimager (BioRad GS-250 Molecular Imager).

RESULTS AND DISCUSSION

Design and synthesis of SBC ODNs

The design paradigm for SBC ODNs is modification of complementary dA-dT or dG-dC bases such that the modified bases form only one hydrogen bond when paired to each other, yet can form two or even three hydrogen bonds when paired to the natural partner. We report the synthesis of a complementary pair of G/C-rich SBC 28mers substituted with deoxyinosine (**dI**) in place of dG and 3-(2′-deoxy-β-D-ribofuranosyl) furano-[2,3-*d*]-pyrimidine-6(5*H*)-one (**dF**) in place of dC. As shown in Figure 2, these modified bases should form two hydrogen bonds, respectively, with dC (2b) or dG (2c), yet only one hydrogen bond with each other (2d). Although the stabilities of the SBC–DNA hybrids might not be as good as DNA–DNA hybrids, the SBC–SBC hybrids would be much less stable, thus enabling the design goals.

The dG analog was simply prepared by removal of the N2 exocyclic amino group of dG to give deoxyinosine (**dI**). This



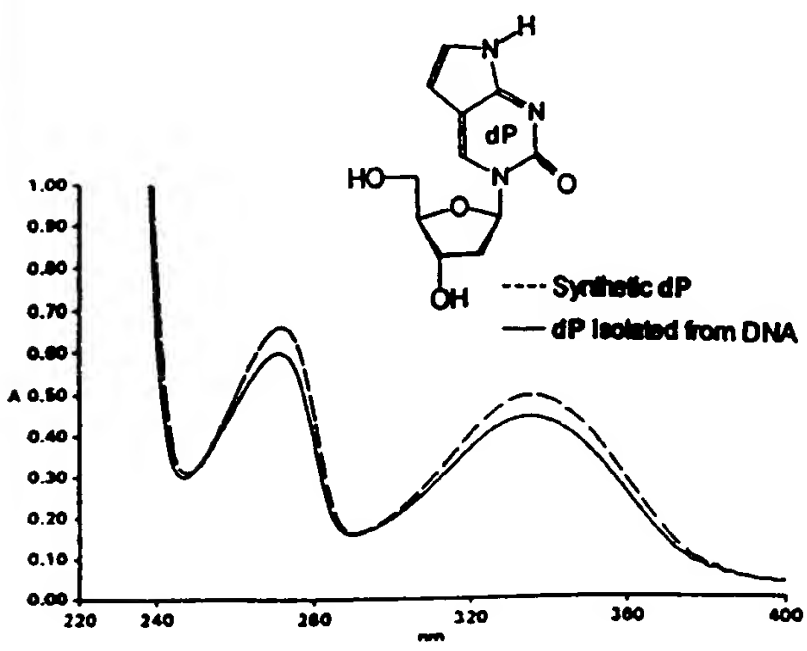


Figure 4. Ultraviolet spectra of **dF** (upper) and **dP** (lower).

nucleoside analog is known to preferentially pair with dC (31–32). The modified dC was designed to have no hydrogen bonding ability at the position equivalent to the N4 exocyclic amino group of dC. We chose the bicyclic nucleoside **dF** to fulfill this role. It was expected to be better than monocyclic dC analogs (33) because of its base stacking ability. The nucleoside **dF** was synthesized by copper (I)-catalyzed cyclization of the known antiviral nucleoside, 5-ethynyl-2′-deoxyuridine (23). Unlike **dF** analogs with a substituent at C2 (23, 34–35), preparation of the desired compound was very sensitive to solvent and reaction conditions; for example, if pyridine was used as a solvent the major product was a dimer of the starting material. **dF** was dimethoxytritylated and converted to its cyanoethoxy phosphoramidite by conventional methods (25).

SBC and unmodified ODNs were synthesized using standard procedures on an ABI-394 DNA synthesizer. Based upon recoveries of purified products, the average coupling yield was 91% for SBC ODNs and 94% for unmodified ODNs. The nucleoside composition of a representative SBC ODN was determined by reverse phase HPLC analysis of an enzyme hydrolysate. As shown in Figure 3, no peak corresponding to authentic **dF** was observed in the nucleoside hydrolysate from the modified ODN. Instead a peak

Table 1. T_m values for native and modified ODNs with dI and dP

Watson: 5′ XTY AXA AXY ATX YYA YYA XXY AAY YAY X 3′
Crick: 3′ YAX TYT TYX TAY XXT XXT YYX TTX XTX Y 5′

Hybrid	Watson X	Watson Y	Crick X	Crick Y	T_M (°C)	T_M Drop per Modified Base Pair
I	C	G	C	G	75.6	0
II	C	I	C	G	55.2	2.04
III	C	G	C	I	62.8	1.83
IV	C	I	C	I	43	1.92
V	P	G	C	G	70.6	0.71
VI	C	G	P	G	72	0.36
VII	P	G	P	G	69.6	0.35
VIII	P	I	C	G	48.2	1.61
IX	C	G	P	I	57.2	1.08
X	P	I	P	I	20.2	3.26

corresponding to 3-(2′-deoxy-β-D-ribofuranosyl)-pyrrolo-[2,3-*d*]-pyrimidine-2(3*H*)-one (**dP**) was observed, suggesting that the **dF** had been converted to **dP** during the treatment with 30% aqueous ammonium hydroxide used in the final step in DNA synthesis.

The **dP** isolated from the modified ODN was identical to authentic **dP** by both UV analysis (Fig. 4) and HPLC coinjection in two gradient solvent systems. The **dF** nucleoside, when treated overnight with 30% aqueous NH_4OH at room temperature, rearranged to a compound (95.4% recovered yield) which had a H^1 NMR identical to that previously reported for **dP** (28). Based on these results we conclude that **dP** was the only dC analog detectable in the ODN hydrolysate and that >90% of the **dF** residues had been converted to **dP** by base treatment. Subsequent attempts to incorporate **dP** directly into ODNs using the phosphoramidite method were not successful due to instability during the iodine oxidation conditions employed in the standard ODN synthesis cycle.

Although **dP** still could hydrogen bond at N1 with the carbonyl group at C6 of dG or **dI**, this interaction should be relatively weak due to suboptimal orientation of the N1 hydrogen (Fig. 2e and f). Previous studies on the base pairing properties of **dP** have shown that it preferentially pairs with dG and that this base pair is slightly less stable than dC-dG (28).

Hybridization properties of SBC ODNs

Table 1 shows the hybridization properties of 28mer ODNs containing **dI** for dG, **dP** for dC, or both. The sequence, taken from pBR322 plasmid, had a G-C content of 60.7%. Introduction of either **dI** or **dP** into one or both strands of the duplex decreased its T_m by 1.8–2.0 or 0.4–0.7°C, respectively, per modified base pair. When only one strand of the hybrid was substituted with both **dI** and **dP**, the T_m dropped by 1.1–1.6°C per modified base pair. These values reflect a slight destabilization attributable to the dG-**dP** base pair and a larger destabilization due to the **dI**-dC base pair. When both strands of the hybrid were substituted with **dI** and **dP**, however, the T_m drop per modified base pair increased significantly to 3.3°C.

Some of the hybrids were analyzed by non-denaturing PAGE (Fig. 5). As shown in Table 1 and Figure 5, the SBC ODNs containing both **dI** for dG and **dP** for dC (Watson in **VIII**, Crick in **IX** and both Watson and Crick in **X**) did not form a stable hybrid with each other at room temperature (hybrid **X**; Fig. 5, lane 8), yet did form stable hybrids with their unmodified complementary

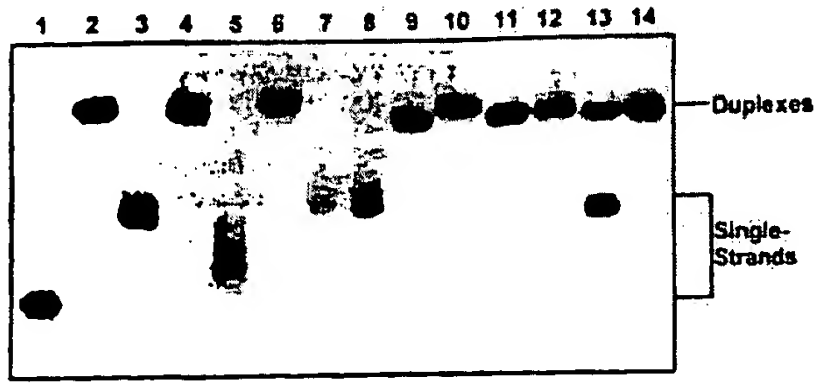


Figure 5. Gel mobility shift analysis of selected ODNs and hybrids listed in Table 1. ^{32}P end-labeled ODNs are denoted by an asterisk (*) and Watson/Crick strands by W/C (see experimental section for details). Single-stranded ODN was analyzed in lanes 1 (*W in I), 3 (*C in IV), 5 (*C in VII), and 7 (*C in X). Complementary ODNs were analyzed in lanes 2 (*W/C in I), 4 (W/*C in IV), 6 (W/*C in VII), 8 (W/*C in X), 9 (*W/C in III), 10 (*W/C in VI), and 11 (*W/C in IX). For lane 12 (W/*C in IX + C in I; molar ratio 2:1:3), the preformed hybrid **IX** was treated at room temperature for 60 min with unmodified Crick. For lane 13 (W in I + C in I + *C in IX; molar ratio 1:1:1), unmodified Watson (1 μM) was mixed simultaneously with unmodified Crick and SBC Crick and incubated at room temperature for 60 min. For lane 14 (W/C in **VIII** plus W/*C in **IX**; molar ratio 2:1:2:1), the hybrid **VIII** was mixed with the hybrid **IX** and incubated at room temperature overnight.

ODN strands (hybrid **IX**; Fig. 5, lane 11). As a result, these ODNs exhibited selective complementary binding. Despite the reduced stability of hybrids formed between SBC and normal ODNs, the normal Watson strand showed no preference for the normal Crick over the SBC Crick strand when equimolar of these three strands were mixed simultaneously at room temperature; about equal amount of duplexes **I** and **IX** were formed (Fig. 5, lane 13). Additionally, there was little, if any, strand displacement or strand exchange when the pre-formed DNA–SBC duplex **IX** was incubated with the normal homolog of the SBC strand or with SBC–DNA duplex **VIII**; not much single-stranded SBC was formed (Fig. 5, lanes 12 and 14). These data clearly demonstrate that the SBC ODNs described above behaved like natural ODNs when hybridized to unmodified complements, yet did not form stable hybrids with themselves.

Strand invasion of a DNA duplex

To determine whether an SBC ODN could strand invade dsDNA, a 17 base pair segment of hybrid **X** was synthesized as a single self-complementary ODN (**XIII**; Fig. 6A). Linking the complementary domains into one ODN was expected to improve the kinetics of strand invasion and the stability of the product. The chimeric SBC ODN had a T_m of 31 °C and hybridized to a partial DNA complement (Watson in **XI**) at room temperature (Fig. 6B, lane 4). The corresponding unmodified ODN (**XII**), derived from hybrid **I**, had a T_m of 80 °C and hybridized poorly to the same DNA complement (Fig. 6B, lane 3). A 48 bp duplex (**XI**) with one end homologous to the self-complementary ODNs was used as a substrate for strand invasion. Annealing was facilitated by the presence of two 5 base long single-stranded overhangs in **XI** which could hybridize to complementary four base long overhangs in the invading ODNs. The duplex can be likened to the stem of a hairpin that might exist in a long ssDNA. After 4 h at room temperature, a 10-fold excess of **XIII** converted 73% of **XI** to a three-way junction compared with 17% for **XII** (Fig. 6B, lanes 6 and 5). Ongoing studies indicate that the self-complementary SBC ODN has a significant kinetic advantage over the unmodified ODN.

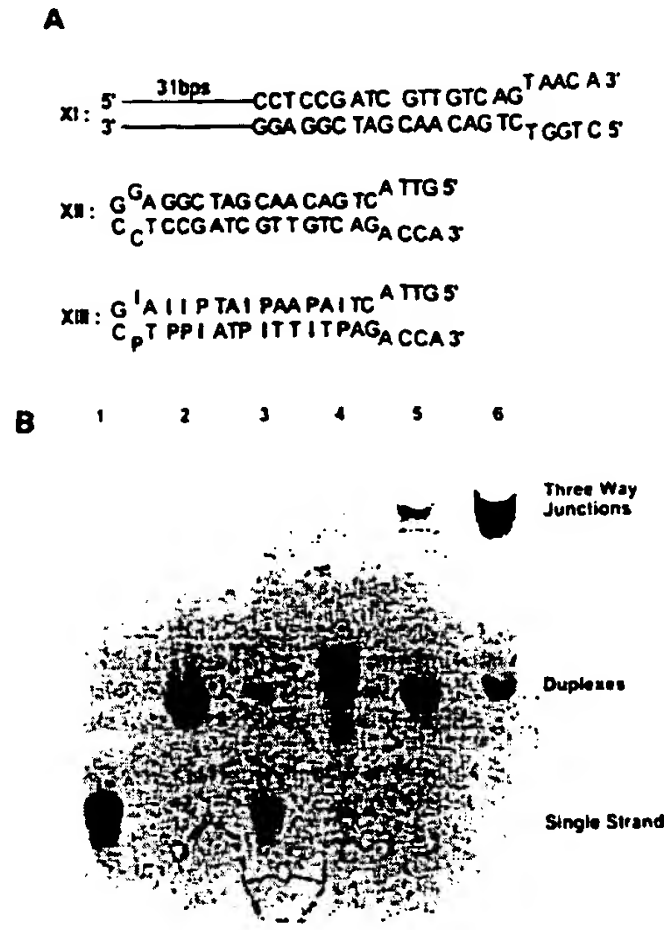


Figure 6. Gel mobility shift analysis of the strand invasion properties of normal and SBC self-complementary ODNs. (**A**) Sequences of the dsDNA duplex (**XI**) and the strand invading normal (**XII**) and SBC (**XIII**) self-complementary ODNs. Employing the convention of Table 1, the upper and lower strands of **XI** are Crick and Watson, respectively. (**B**) Gel mobility shift analysis of strand invasion. Reactions were incubated 4 h at room temperature in TNM buffer prior to electrophoresis. Unless otherwise indicated, hybrids were formed by mixing the labeled ODN (0.1 μM) with a 2-fold molar excess of cold complementary ODN. Single-strand (*W in **XI**) and double-strand (*W/C in **XI**) standards were run in lanes 1 and 2. Hybridization reactions between free *W strand of duplex **XI** and self-complementary ODNs **XII** or **XIII** (molar ratio 1:10) were analyzed in lanes 3 and 4. Strand invasion reactions between duplex **XI** (*W/C) and self-complementary ODNs **XII** or **XIII** (molar ratio 1:10) were analyzed in lanes 5 and 6.

Conclusions

Based on thermal denaturation and non-denaturing gel mobility shift assays, we have designed and synthesized for the first time modified ODNs which exhibit selective complementary hybridization. A self-complementary version of one of these paired ODNs strand invaded a homologous double-stranded DNA better than the corresponding unmodified ODN. The possible diagnostic and therapeutic uses of these ODNs are being explored. Efforts to improve the hybridization properties of SBC ODNs including the modification of dA and dT are also underway.

ACKNOWLEDGEMENTS

We thank Drs I. V. Kutyavin, A. Gall, V. Gorn and E. Lukhtanov for helpful discussions. We thank Mr D. Adams and Ms A. Yang for technical contributions.

REFERENCES

1 Cheng,S., Van Houten,B., Gamper,H.B., Sancar,A. and Hearst,J.E. (1988) *J. Biol. Chem.*, **263**, 15110–15117.
2 Hélène,C. (1993) In Crooke,S.T. and Lebleu,B. (eds) *Antisense Research and Applications*. CRC Press, Boca Raton, pp. 375–385.
3 Iyer,M., Norton,J.C. and Corey,D.R. (1995) *J. Biol. Chem.*, **270**, 14712–14717.
4 Gamper,H.B., Cimino,G.D. and Hearst,J.E. (1987) *J. Mol. Biol.*, **197**, 349–362.
5 Chastain,M. and Tinoco,I.,Jr. (1993) In Crooke,S.T. and Lebleu,B. (eds) *Antisense Research and Applications*. CRC Press, Boca Raton, pp. 55–66.
6 Sproat,B.S. and Lamond,A.I. (1993) In Crooke,S.T. and Lebleu,B. (eds) *Antisense Research and Applications*. CRC Press, Boca Raton, pp. 352–262.
7 Monia,B.P., Lesnik,E.A., Gonzalez,C., Lima,W.F., McGee,D., Guinosso,C.J., Kawasaki,A.M., Cook,P.D. and Freier,S.M. (1993) *J. Biol. Chem.*, **268**, 14514–14522.
8 Gryaznov,S. and Chen,J.-K. (1994) *J. Am. Chem. Soc.*, **116**, 3143–3144.
9 Nielsen,P.E., Egholm,M. and Buchardt,O. (1994) *Bioconjugate Chem.*, **5**, 3–7.
10 Lamm,G.M., Blencowe,B.J., Sproat,B.S., Iribarren,A.M., Ryder,U. and Lamond,A.I. (1991) *Nucleic Acids Res.*, **19**, 3193–3198.
Wagner,R.W., Matteucci,M.D., Lewis,J.G., Gutierrez,A.J., Moulds C. and Froehler,B.C. (1993) *Science*, **260**, 1510–1513.
12 Asseline,U., Delarue,M., Lancelot,G., Toulme,F., Thuong,N.T., Montenay-Garestier,T. and Hélène,C. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3297–3301.
13 Wang,S. and Kool,E.T. (1994) *J. Am. Chem. Soc.*, **116**, 8857–8858.
14 Lima,W.F., Monia,B.P., Ecker,D.J. and Freier,S.M. (1992) *Biochemistry*, **31**, 12055–12061.
15 Ecker,D.J., Vickers,T.A., Bruice,T.W., Freier,S.M., Jenison,R.D., Manoharan,M. and Zounes,M. (1992) *Science*, **257**, 958–961.
16 Francois,J.-C., Thuong,N.T. and Hélène,C. (1994) *Nucleic Acids Res.*, **22**, 3943–3950.
17 Brossalina,E. and Toulme,J.-J. (1993) *J. Am. Chem. Soc.*, **115**, 796–797.
18 Francois,J.C. and Hélène,C. (1995) *Biochemistry*, **34**, 65–72.
19 Kutyavin,I.V., Podyminogin,M.A., Bazhina,Y.N., Fedorova,O.S., Knorre,D.G., Levina,A.S., Mamayev,S.V. and Zarytova,V.F. (1988) *FEBS Lett.*, **238**, 35–38.
20 Richardson,P.L. and Schepartz,A. (1991) *J. Am. Chem. Soc.*, **113**, 5109–5111.
21 Jayasena,V.K. and Johnston,B.H. (1993) *J. Mol. Biol.*, **230**, 1015–1024.
22 Sena,E.P. and Zarling,D.A. (1993) *Nature Genetics*, **3**, 365–372.
23 Robins,M.J. and Barr,P.J. (1983) *J. Org. Chem.*, **48**, 1854–1862.
24 Kumar,R., Knaus,E.E. and Wiebe,L.I. (1991) *J. Heterocyclic Chem.*, **28**, 1917–1925.
25 Sinha,N.D., Biernat,J., McManus,J. and Koster,H. (1984) *Nucleic Acids Res.*, **12**, 4539–4557.
26 Hopkins,P.B., Millard,J.T., Woo,J., Weidner,M.F., Kirchner,J.J., Sigurdsson,S.T. and Raucher,S. (1991) *Tetrahedron*, **47**, 2475–2489.
27 Woo,J., Sigurdsson,S.T. and Hopkins,P.B. (1993) *J. Am. Chem. Soc.*, **115**, 3407–3415.
28 Inoue,H., Imura,A., and Ohtsuka,E. (1987) *Nippon Kagaku Kaishi*, **7**, 1214–1220.
29 Albergo,D.D., Markey,L.A., Breslauer,K.J. and Turner,D.H. (1981) *Biochemistry*, **20**, 1409–1416.
30 Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 125–126.
31 Case-Green,S.C. and Southern,E.M. (1994) *Nucleic Acids Res.*, **22**, 131–136.
32 Martin,F.H., Castro,M.M., Aboul-ela,F. and Tinoco,I. (1985) *Nucleic Acids Res.*, **13**, 8927–8938.
33 Gildea,B. and McLaughlin,L.W. (1989) *Nucleic Acids Res.*, **17**, 2261–2281.
34 Robins,M.J., Vinayak,R.S. and Wood,S.G. (1990) *Tetrahedron Lett.*, **31**, 3731–3734.
35 Crisp,G.T. and Flynn,B.L. (1993) *J. Org. Chem.*, **58**, 6614–6619.

US005912340A

United States Patent [19]

Kutyavin et al.

[11] **Patent Number: 5,912,340**

[45] **Date of Patent: Jun. 15, 1999**

[54] **SELECTIVE BINDING COMPLEMENTARY OLIGONUCLEOTIDES**

[75] Inventors: **Igor V. Kutyavin**, Bothell; **Jinsuk Woo**, Lynnwoode; **Eugeny A. Lukhtanov**; **Rich B. Meyer, Jr.**, both of Bothell; **Howard B. Gamper**, Woodinville, all of Wash.

[73] Assignee: **Epoch Pharmaceuticals, Inc.**, Bothell, Wash.

[21] Appl. No.: **08/539,097**

[22] Filed: **Oct. 4, 1995**

[51] **Int. Cl.**[6] **C07H 21/04**

[52] **U.S. Cl.** **536/24.5**; 536/26.6; 536/24.3

[58] **Field of Search** 536/24.5, 26.6, 536/24.3

[56] **References Cited**

FOREIGN PATENT DOCUMENTS

WO93/03736	3/1993	WIPO	A61K 31/70
95/05391	2/1995	WIPO .	
95/14707	6/1995	WIPO	C07H 19/16

OTHER PUBLICATIONS

Scheit et al. Studia Biophysica vol. 55, No. 1, 1976. pp. 21–27.
Inoue et al. Chemical Abstracts vol. 108, No. 21, 1988. p. 752, col. 1.
Database WPI, AN 87–352165[50] (see abstract) & Patent Abstracts of Japan, vol. 012, No. 139 (see abstract).
Case–Green et al. Nucleic Acids Research, vol. 22, No. 2, 1994. pp. 131–136.
Martin et al. Nucleic Acids Research, vol. 13, No. 24, 1985. pp. 8927–8938.
Chollet et al. Nucleic Acids Research, vol. 16, No. 1, 1988. pp. 305–317.
Kuimelis et al. Nucleic Acids Research, vol. 22, No. 8, 1994. pp. 1429–1436.
Ishikawa et al. Chemical Abstracts, vol. 116, No. 13, 1992. pp. 949. col. 2.
Newman et al. Biochemistry, vol. 29, No. 42, 1990. pp. 9891–9901.
Richardson et al. Journal of the American Chemical Society, vol. 113, No. 13, 1991. pp. 5109–5111.
Woo J et al. Nucleic Acids Research, vol. 24, No. 13, 1996. pp. 2470–2475.2.
Strobel, S.A., et al. (1991) *Science*, 254:1639.
Weinstock, P., et al. (1990) *Nucl. Acids Res.*, 18:4207.
Roca, A.I., et al. (1990) *Rev. Biochem. Mol. Biol.*, 25:415.
Robins et al.(1982) Can. J. Chem., 60:554.
Robins et al.(1983) *J. Org. Chem.*, 48:1854.
Meyer, et al., (1989) *J. Am. Chem. Soc.*, 111:8517.
Kobayashi (1973) *Chem. Phar. Bull.*, 21:941–951.
Sonveaux (1986) Bioorganic Chemistry, 14:274–325.
Jones (1984) "Oligonucleotide Synthesis, a Practical Approach", M.J. Gait, Ed., IRL Press, P23–34.
Langer et al. (1981) *Proc. Natl. Acad. Sci. USA*, 78:6633–6637. "Nucleic Acid Hybridisation, a Practical Approach", Hames and Higgins, Eds., IRL Press, 1985.
Gall and Pardue, (1969) *Proc. Natl. Acad. Sci., U.S.A.*, 63:378–383.
John et al., (1969) Nature, 223:582–587.
"Physical Biochemistry", Freifelder, D., W.H. Freeman & Co., 1982, pp. 537–542.
Tijssen, P., (1985) "Practice and theory of Enzyme Immunoassays, Laboratory Techniques in Biochemistry and Molecular Biology", Burdon, R.H., van Knippenberg, P.H., Eds, Elsevier, pp. 9–20.
Connolly, et al., (1989) *Nucleic Acids Res.*,17:4957–4974.
Fathi, et al. (1990) *Tetrahedron Lett.*, 31:319–322.
Sinha, et al. (1984) Nucleic Acids Research, 12:4539.
Alul, et al. (1991) *Nucleic Acids Res.*, 19:1527–1532.
Atkinson, T. and Smith, M. (1984) "Oligonucleotide Synthesis, a Practical Approach", M. Gait, Ed., IRL Press, Washington, D.C., pp. 35–81.

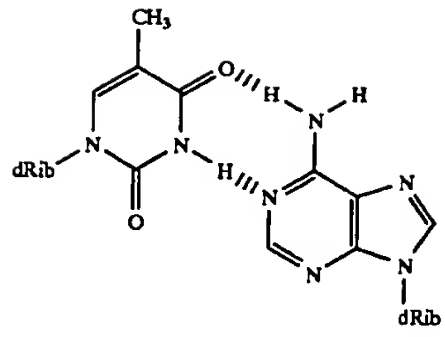
Primary Examiner—Scott W. Houtteman
Attorney, Agent, or Firm—Klein & Szekeres, LLP

[57] **ABSTRACT**

In a matched pair of oligonucleotides (ODNS) each member of the pair is complementary or substantially complementary in the Watson Crick sense to a target sequence of duplex nucleic acid where the two strands of the target sequence are themselves complementary to one another. The ODNs include modified bases of such nature that the modified base forms a stable hydrogen bonded base pair with the natural partner base, but does not form a stable hydrogen bonded base pair with its modified partner. This is accomplished when in a hybridized structure the modified base is capable of forming two or more hydrogen bonds with its natural complementary base, but only one hydrogen bond with its modified partner. Due to the lack of stable hydrogen bonding with each other, the matched pair of oligonucleotides have a melting temperature under physiological or substantially physiological conditions of approximately 40° C. or less. However each of the matched ODN pair of the invention forms a substantially stable hybrid with the target sequence in each strand of the duplex nucleic acid. The hybrids of target duplex nucleic acids formed with the ODN pairs of the invention are useful for gene mapping and in diagnostic and therapeutic applications.

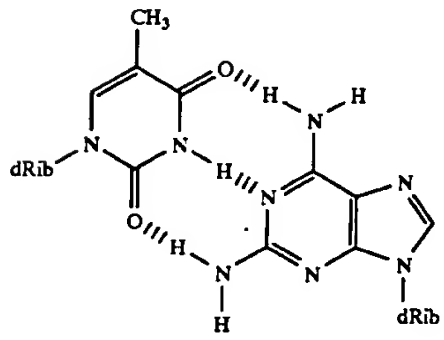
26 Claims, No Drawings

1a



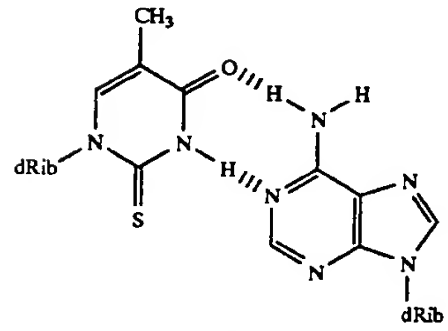
T:A

1b

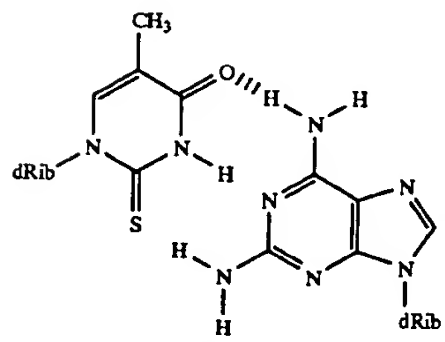


T:2-amA

2a

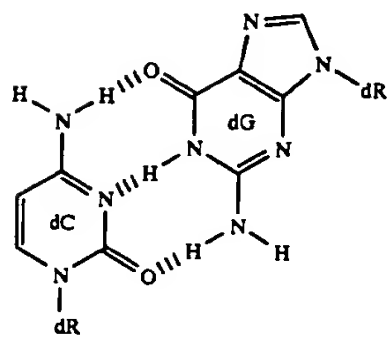


2-sT:A



2-sT:2-amA

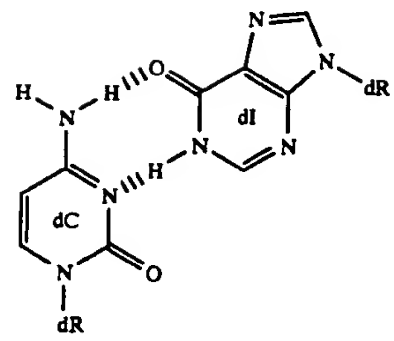
2b



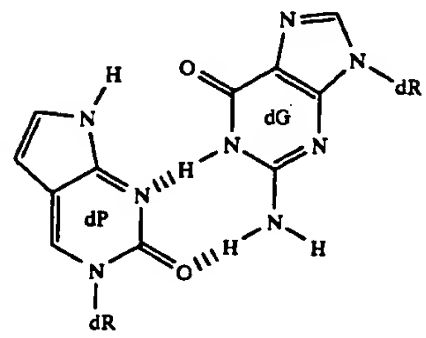
3a

-continued

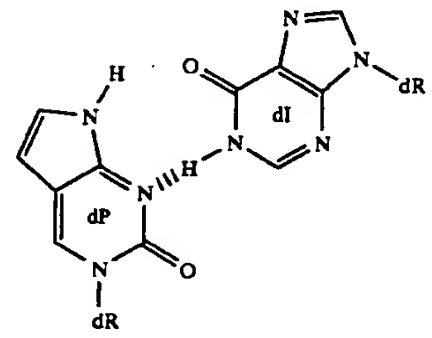
3b



4a



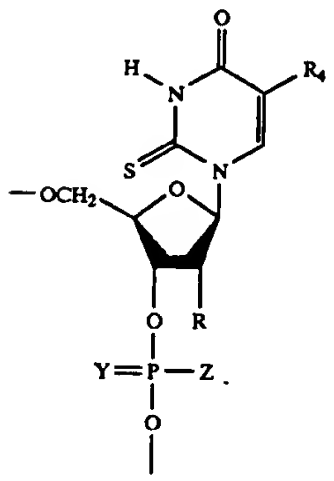
4b



A sufficient number of the modified SBC nucleotides are incorporated such that complementary positions in both SBC ODNS are modified into a matched pair of SBC ODNs of the present invention so that the pair of the matched set does not form a stable hybrid; in other words under physiological conditions it has a melting temperature of approximately 40° C. or less. It is not necessary to replace each natural nucleotide of the ODN with a modified SBC nucleotide in order to accomplish this. Both members of the matched pair are however complementary to a target sequence in double stranded or duplex nucleic acid, where the two strands or parts of the target duplex are themselves complementary or substantially complementary to one another. As it is described in more detail below, an important use of the SBC ODNs of the present invention is hybridization with secondary structure of mRNA wherein the mRNA itself forms a duplex, such as in hairpin loops. It is known that secondary structure of mRNA and ribosomal RNA do not have two strands in the strict sense of that term. Nevertheless, unless the context otherwise indicates, in the present description the terminology "two strands" of double stranded nucleic acids also refers to the two complementary portions of duplex mRNA or of duplex ribosomal RNA as well. The general concept of double stranded DNA and of secondary structure in mRNA and ribosomal RNA is covered in this description by the term "duplex nucleic acid". The term "RNA" can apply to any functional RNA in living organisms, such as messenger, transfer, ribosomal, small nuclear, guide, genomic, etc. RNA.

R_4 is H, C_{1-6} alkyl, C_{1-6} alkenyl, C_{1-6} alkynyl, or optionally the 5-position of the pyrimidine serves as point of attachment for a cross-linking function, or a reporter group as described below. A preferred embodiment of the SBC nucleotide T' has 2-thio-4-oxo-5-methylpyrimidine (2-thiothymine) as the base, as shown in Formula 2b. The latter nucleotide is abbreviated as 2-sT or d 2-sT as applicable.

Formula 8

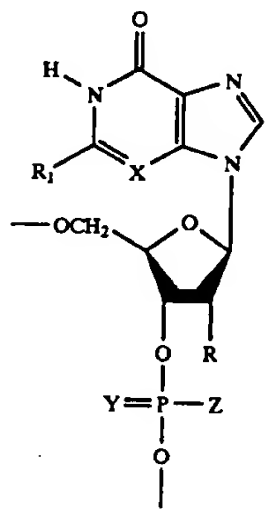


A general structure for a preferred class of the modified G analog, G', within the scope of the invention and shown as a 3'-phosphate (or phosphorothioate) incorporated into the SBC ODN, is provided by Formulas 9, 10 and 11, wherein

R_1 is H, C_{1-4} alkyl, C_{1-4} alkoxy, C_{1-4} alkylthio, F or NHR_3 where R_3 is defined as above,

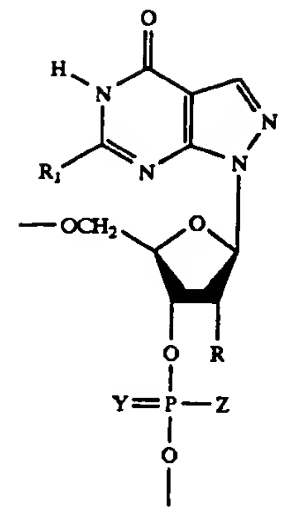
X, Y, Z and R are defined as above, and the 8 position of the purine, the 3 position of the pyrrazolopyrimidine or the 5 position of the pyrrolopyrimidine optionally serve as point of attachment for a cross-linking agent, or reporter group as described below. A preferred embodiment of the SBC nucleotide G' has 6-oxo-purine (hypoxanthine) as the base, as shown in Formula 3b. The latter nucleotide is abbreviated as I or dI as applicable.

Formula 9

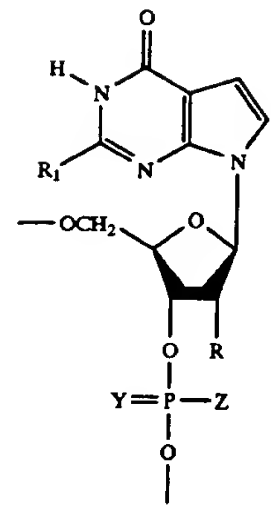


-continued

Formula 10



Formula 11



A general structure for a preferred class of the modified C analog, C', within the scope of the invention and shown as a 3'-phosphate (or phosphorothioate) incorporated into the SBC ODN, is provided by Formulas 12 and 13, wherein

Y, Z, R and R_4 are defined as above, or optionally the 5-position of the pyrimidine serves as point of attachment for a cross-linking function, or a reporter group as described below;

Z_1 is O or NH, and

R_5 is H or C_{1-4} alkyl.

A preferred embodiment of the SBC nucleotide C' has pyrrolo-[2,3-d]pyrimidine-2(3H)-one as the base, as shown in Formula 4b. The latter nucleotide is abbreviated as P or dP as applicable.

$-(CH_2)_q-NH-CO-(CH_2)_m-(X^*)_n-N(R_1)-(CH_2)_p-L$ and

$-(CH_2)_{q'}-O-(CH_2)_{q''}-NH-CO-(CH_2)_m-(X^*)_n-N(R_1)-(CH_2)_p-L$

where q, m and L are defined as above, q' is 3 to 7 inclusive, q" is 1 to 7 inclusive, X* is phenyl or simple substituted phenyl (such as chloro, bromo, lower alkyl or lower alkoxy substituted phenyl), n is 0 or 1, p is an integer from 1 to 6, and R_1 is H, lower alkyl or $(CH_2)_p$—L. Preferably p is 2. Those skilled in the art will recognize that the structure $-N(R_1)-(CH_2)_2-L$ describes a "nitrogen mustard", which is a class of potent alkylating agents. Particularly preferred within this class of BBC ODNs of the invention are those where the cross-linking agent includes the functionality $-N(R_1)-(CH_2)_2-L$ where L is halogen, preferably chlorine; and even more preferred within this class are those modified SBC ODNs where the cross linking agent includes the grouping $-N-[(CH_2)_2-L]_2$ (a "bifunctional" N-mustard).

A particularly preferred partial structure of the cross linking agent includes the grouping

$-CO-(CH_2)_3-C_6H_4-N-[(CH_2)_2Cl]_2$.

In a particularly preferred embodiment the just-noted cross-linking group is attached to an n-hexylamine bearing tail at the 5' and 3' ends of the SBC ODN in accordance with the following structure:

$R'-O-(CH_2)_6-NH-CO-(CH_2)_3-C_6H_4-N-[(CH_2)_2Cl]_2$

where R' signifies the terminal 5' or 3'—phosphate group of the SBC ODN.

Other examples for the A*—L group, particularly when attached to a heterocyclic base in the oligonucleotide (such as to the 5-position of 2'-deoxyuridine) are 3-iodoacetamidopropyl, 3-(4-bromobutyramido)propyl, 4-iodoacetamidobutyl and 4-(4-bromobutyramido)butyl groups.

In accordance with other preferred embodiments, the cross-linking functionality is covalently linked to the heterocyclic base, for example to the uracil moiety of a 2'-deoxyuridylic acid building block of the SBC ODN. The linkage can occur through the intermediacy of an amino group, that is, the "arm-leaving group combination" (A*—L) may be attached to a 5-amino-2'-deoxyuridylic acid building unit of the SBC ODN. In still other preferred embodiments the "arm-leaving group combination" (A*—L) is attached to the 5-position of the 2'-deoxyuridylic acid building unit of the SBC ODN by a carbon-to-carbon bond. Generally speaking, 5-substituted-2'-deoxyuridines can be obtained by an adaptation of the general procedure of Robins et al. (*Can. J. Chem.*, 60:554 (1982); *J. Org. Chem.*, 48:1854 (1983)), as shown in Reaction Scheme 1. In accordance with this adaptation, the palladium-mediated coupling of a substituted 1-alkyne to 5-iodo-2'-deoxyuridine gives an acetylene-coupled product. The acetylenic dUrd analog is reduced, with Raney nickel for example, to give the saturated compound, which is then used for direct conversion to a reagent for use on an automated DNA synthesizer, as described below. In Reaction Scheme 1, q is defined as above, and Y' is either Y* (as defined above) or is a suitable protected derivative of Y*. Y' can also be defined as a group which terminates in a suitably protected nucleophilic function, such as a protected amine. Examples of reagents which can be coupled to 5-iodo-2'-deoxyuridine in accordance with this scheme are

$HC\equiv CCH_2OCH_2CH_2N(CO)_2C_6H_4$ (phtalimidoethoxypropyne),

$HC\equiv CCH_2OCH_2CH_2NHCOCF_3$ (trifluoroacetamidoethoxypropyne),

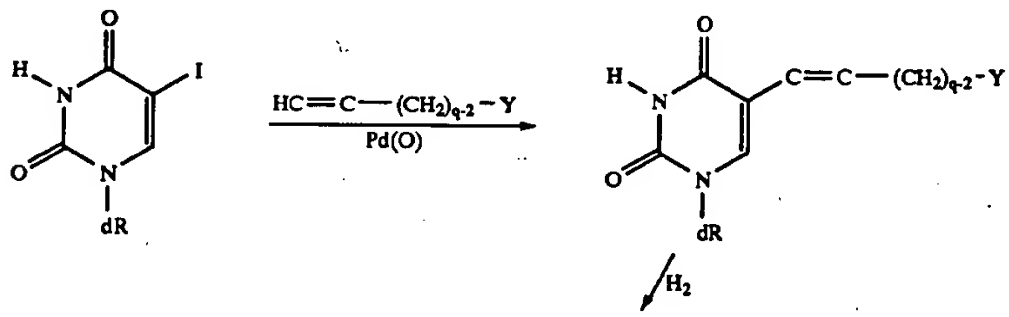
$HC\equiv CCH_2N(CO)_2C_6H_4$ (phtalimidopropyne) and

$HC\equiv CCH_2NHCOCF_3$ (trifluoroacetamidopropyne),

In these examples the nucleosides which are obtained in this scheme are incorporated into the desired SBC ODN, and the alkylating portion of the cross-linking agent is attached to the terminal amino group of "Y'" only after removal of the respective phtalic or trifluoroacetyl blocking groups.

Another particularly preferred example of an "arm-leaving group combination" (A*—L) is attachment of a nitrogen-mustard type alkylating agent (or other alkylating agent) to the amino function of a 5-(3-aminopropyl)-2'-deoxyuridine building unit of the SBC ODN. The appropriate nucleotide building unit for ODN synthesis which includes the 5-(3-aminopropyl)-2'-deoxyuridine nucleoside moiety can be obtained in analogy to Reaction Scheme 1, and in accordance with the teaching of Meyer et al., J. Am. Chem. Soc. 1989, 111, 8517. In this particularly preferred embodiment the nucleotide having the 5-(3-aminopropyl)-2'-deoxyuridine moiety is incorporated into the SBC ODN by routine synthesis, and the cross-linking function is introduced by reacting the BBC ODN with an activated form of a "nitrogen mustard", such as 2,3,5,6-tetrafluorophenyl-4'-[bis(2-chloroethyl)amino]phenylbutyrate (Chlorambucil 2,3,5,6-tetrafluorophenyl ester; chlorambucil itself is commercially available).

Reaction Scheme 1



A moiety containing the leaving group, such as a haloacyl group ($CO—CH_2—L$ where L is halogen for example I) or $—CO—(CH_2)_m—(X^*)_n—N(R_1)—(CH_2)_p—L$ group (even more preferably a $CO—(CH_2)_3—C_6H_4—N—[CH_2CH_2Cl]_2$) may be added to the aminoalkyl or like groups ($—(CH_2)_q—Y^*$) following incorporation into oligonucleotides and removal of any blocking groups. For example, addition of an α-haloacetamide may be verified by a changed mobility of the modified compound on HPLC, corresponding to the removal of the positive charge of the amino group, and by subsequent readdition of a positive charge by reaction with 2-aminoethanethiol to give a derivative with reverse phase HPLC mobility similar to the original aminoalkyl-oligonucleotide.

In the situations where the cross linking agent ($A^*—L$ moiety) is attached to the 3' or 5' terminus of the oligonucleotide, for example by an alkylamine linkage of the formula $—(CH_2)_q—Y^*$ (Y^* terminating in an amine), the oligonucleotide synthesis may be performed to first yield the oligonucleotide with said aminoalkyl tail, to which then an alkylating moiety, such as the above-noted haloacylgroup ($CO—CH_2—L$) or $—CO—(CH_2)_m—(X^*)_n—N(R_1)—(CH_2)_p—L$ is introduced.

SBC ODNs bearing a reporter group, lipophilic group or tail

As is known in the art a "reporter group" can be broadly defined as a group that is incorporated in, or is attached to an ODN and which renders detection or isolation of the ODN possible by application of some analytical, physical, chemical or biochemical method. Generally speaking reporter groups are attached to ODNs hen the ODNs are used as probes. In terms of attaching reporter groups to ODNs in the general sense, the art is well developed and is recited here only in a summary fashion. The SBC ODNs of the present invention having a reporter group (such as a radiactive label) attached, can be utilized substantially in accordance with state-of-the-art hybridization technology, to detect specific target sequences in duplex regions of nucleic acids. The advantage of the SBC ODNs of the present invention, as compared to the prior art, is that the SBC ODN of the present invention can effectively invade and bind to the duplex nucleic acid sequence.

Thus, probes may be labeled by any one of several methods typically used in the art. A common method of detection is the use of autoradiography with ^3H, ^{125}I, ^{35}S, ^{14}C, or 32P labeled probes or the like. Other reporter groups include ligands which bind to antibodies labeled with fluorophores, chemiluminescent agents, and enzymes. Alternatively, probes can be conjugated directly with labels such as fluorophores, chemiluminescent agents, enzymes and enzyme substrates. Alternatively, the same components may be indirectly bonded through a ligand-antiligand complex, such as antibodies reactive with a ligand conjugated with label. The choice of label depends on sensitivity required, ease of conjugation with the probe, stability requirements, and available instrumentation.

The choice of label dictates the manner in which the label is incorporated into the probe. Radioactive probes are typically made using commercially available nucleotides containing the desired radioactive isotope. The radioactive nucleotides can be incorporated into probes, for example, by using DNA synthesizers, by nick-translation, by tailing of radioactive bases in the 3' end of probes with terminal transferase or the 5'-end with a polynucleotide kinase.

Non-radioactive probes can be labeled directly with a signal (e.g., fluorophore, chemiluminescent agent or enzyme) or labeled indirectly by conjugation with a ligand. For example, a ligand molecule is covalently bound to the probe. This ligand then binds to a receptor molecule which is either inherently detectable or covalently bound to a detectable signal, such as an enzyme or photoreactive compound. Ligands and antiligands may be varied widely. Where a ligand has a natural "antiligand", namely ligands such as biotin, thyroxine, and cortisol, it can be used in conjunction with its labeled, naturally occurring antiligand. Alternatively, any haptenic or antigenic compound can be used in combination with a suitably labeled antibody. A preferred labeling method utilizes biotin-labeled analogs of oligonucleotides, as disclosed in Langer et al., *Proc. Natl. Acad. Sci. USA* 78:6633–6637 (1981), which is incorporated herein by reference.

Enzymes of interest as reporter groups will primarily be hydrolases, particularly phosphatases, esterases, ureases and glycosidases, or oxidoreductases, particularly peroxidases. Fluorescent compounds include fluorescein and its derivatives, rhodamine and its derivatives, dansyl, umbelliferone, rare earths, etc. Chemiluminescers include luciferin, acridinium esters and 2,3-dihydrophthalazinediones, e.g., luminol. A further description of reporter groups and specific examples thereof can be found in U.S. Pat. No. 5,419,966, the specification of which is expressly incorporated herein by reference.

The specific hybridization conditions are not critical and will vary in accordance with the investigator's preferences and needs. The particular hybridization technique is not essential to the invention. Hybridization techniques are generally described in "Nucleic Acid Hybridization, A Practical Approach", Hames and Higgins, Eds., IRL Press, 1985; Gall and Pardue, *Proc. Natl. Acad. Sci., U.S.A.*, 63:378–383 (1969); and John et al., *Nature*, 223:582–587 (1969). As improvements are made in hybridization techniques, they can readily be applied.

The amount of labeled probe which is present in the hybridization solution may vary widely. Generally, substantial excess of probe over the stoichiometric amount of the target duplex nucleic acid will be employed to enhance the rate of binding of the probe to the target sequence.

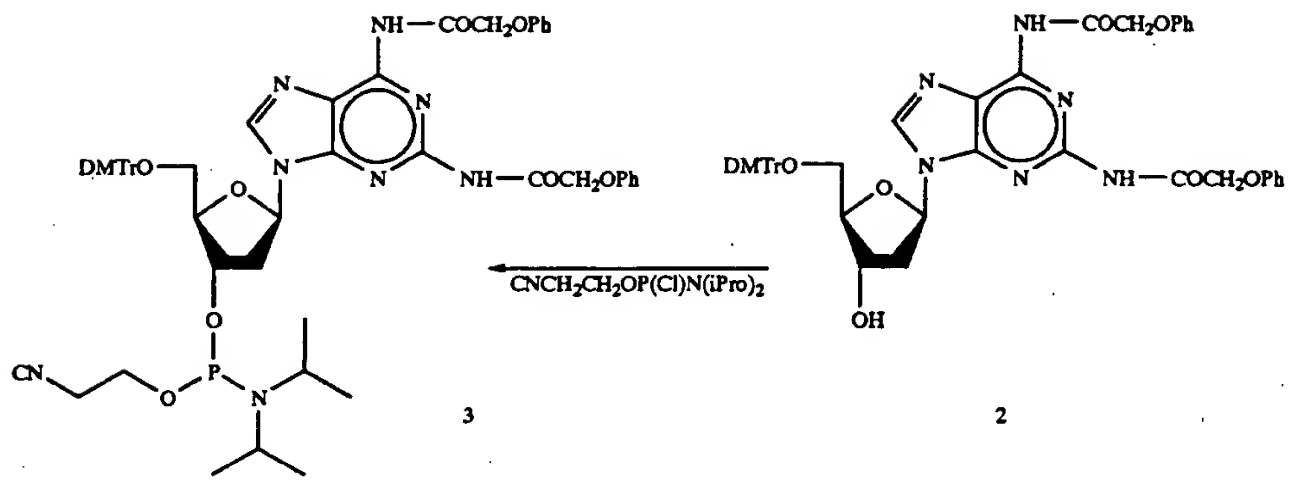
After hybridization at a temperature and time period appropriate for the particular hybridization solution used, the glass, plastic, or filter support to which the probe-target hybrid is attached is introduced into a wash solution typically containing similar reagents as provided in the hybridization solution. Either the hybridization or the wash medium can be stringent. After appropriate stringent washing, the correct hybridization complex may now be detected in accordance with the nature of the label.

The probe may be conjugated directly with the label. For example, where the label is radioactive, the support surface with associated hybridization complex substrate is exposed to X-ray film. Where the label is fluorescent, the sample is detected by first irradiating it with light of a particular wavelength. The sample absorbs this light and then emits light of a different wavelength which is picked up by a detector ("Physical Biochemistry", Freifelder, D., W. H. Freeman & Co., 1982, pp. 537–542). Where the label is an enzyme, the sample is detected by incubation with an appropriate substrate for the enzyme. The signal generated may be a colored precipitate, a colored or fluorescent soluble material, or photons generated by bioluminescence or chemiluminescence. The preferred label for dipstick assays generates a colored precipitate to indicate a positive reading. For example, alkaline phosphatase will dephosphorylate indoxyl phosphate which then will participate in a reduction reaction to convert tetrazolium salts to highly colored and insoluble formazans.

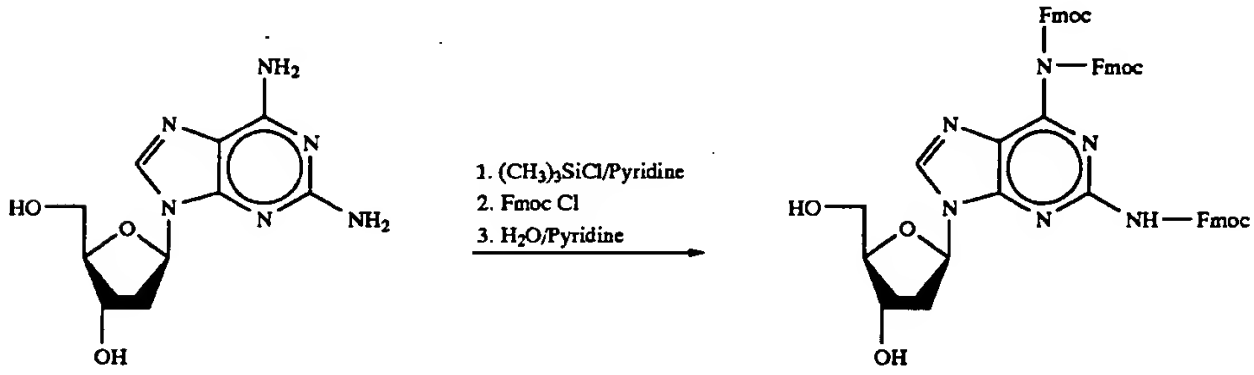
Detection of a hybridization complex may require the binding of a signal generating complex to a duplex of target and probe polynucleotides or nucleic acids. Typically, such binding occurs through ligand and antiligand interactions as between a ligand-conjugated probe and an antiligand conjugated with a signal.

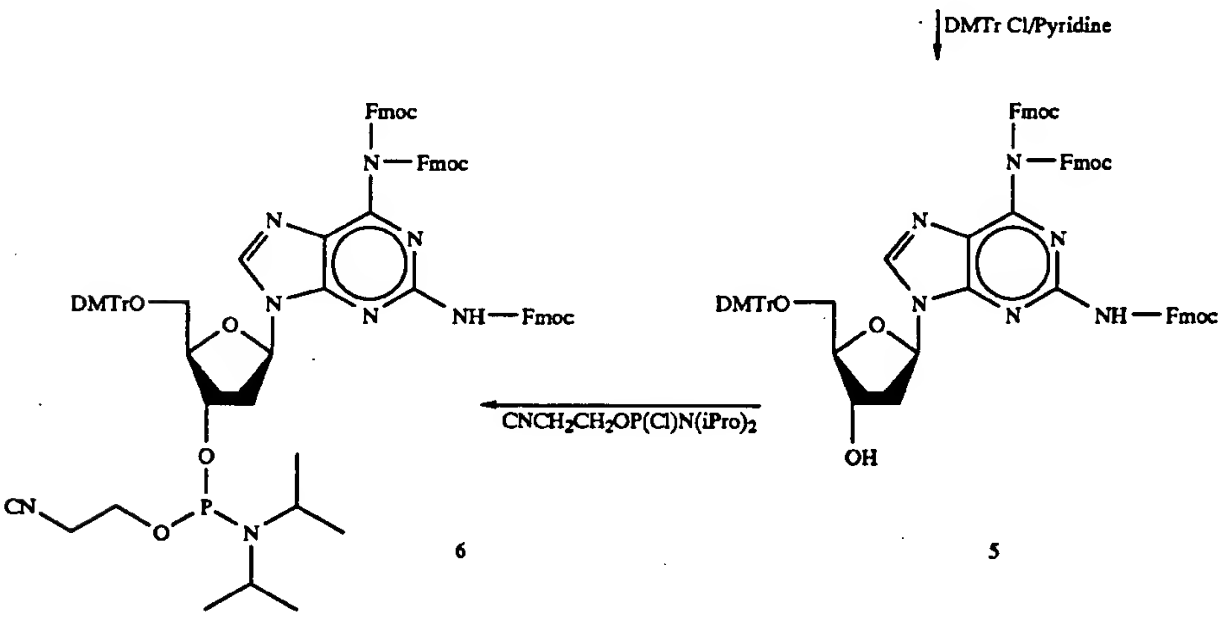
The label may also allow indirect detection of the hybridization complex. For example, where the label is a hapten or

-continued



Reaction Scheme 3





The nucleotide moiety shown in Formula 4b is obtained by the method illustrated in Reaction Scheme 4 which substantially follows known chemical literature. First the "furan" analog deoxyribofuranoside, namely 3-(2'-deoxy-β-D-ribofuranosyl)furano-[2,3-d]pyrimidine-6(5H)-one (Compound 11) is synthesized by copper (I)-catalyzed cyclization from the known antiviral nucleoside 5-ethynyl-2'-deoxyuridine (Compound 10), substantially as in the literature procedure of Robins et al. J. Org. Chem. 1983, 48, 1854. This compound is dimethoxytritylated and converted into the corresponding cyanoethoxy phosphoramidite (Compound 12) suitable as a reagent for ODN synthesis, substantially by conventional literature methods (see Sinha et al. Nucleic Acids Research. 1984, 12, 4539). The SBC ODNs of the present invention are then constructed on a solid support. The final step of treating the SBC ODN with ammonia to remove protecting groups, converts the furano-[2,3-d]pyrimidine-6(5H)-one base into the pyrrolo-[2,3-d]

not with the complementary SBC ODN. Thus, members of a matched pair of SBC ODNs were found to form stable hybrids with their respective natural complementary targets, but not with each other. Table 1 below indicates the melting temperatures observed under the conditions indicated in the table, and also the calculated decrease (drop) in melting temperature per modified base pair.

their unmodified complementary strands, while they do not form stable hybrids with themselves.

Table 2 refers to a complementary pair of 20-mer oligodeoxyribonucleotides (ODN V and ODN VI) which are hybridized under substantially physiological conditions (0.2M NaCl, 0.01M Na_2HPO_4, 0.1 mM EDTA, pH7.0, ODN concentration=4×10^{-7}M). The ODNs designated in Table 2

TABLE 1

Table 1. Tm Values for Native and Modified ODNs with dI and dP

Watson:	5'	XTY	AXA	AXY	ATX	YYA	YYA	XXY	AAY	YAY	X 3'
Crick:	3'	YAX	TYT	TYX	TAY	XXT	XXT	YYX	TTX	XTX	Y 5'

		Watson		Crick				
	Hybrid	X	Y	X	Y		Tm (° C.)[a]	Tm Drop per Modified Base Pair
Sequence ID NO: 1	I	C	G	C	G		75.6	0
Sequence ID NO: 3	II	P	I	C	G	Sequence ID NO: 2	48.2	1.61
Sequence ID NO: 7	III	C	G	P	I	Sequence ID NO: 4	57.2	1.08
Sequence ID NO: 8	IV	P	I	P	I		20.2	3.26

[a]10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 50 mM NaCl, 10 mM $MgCl_2$

In Table 1 the 28-mer ODN is a sequence taken from pBR 322 plasmid. Hybrid 1 is formed from complementary oligodeoxynucleotides wherein X and Y are natural dC and dG residues in both ODNs. Thus, Hybrid 1 provides a reference, to which other hybrids formed of modified SBC ODNs can be compared. The pair of SBC ODNs shown as Hybrid IV in Table 1 comprises two 28-mer sequences where each of the natural dG and dC nucleotides is replaced with dI and dP, respectively. Hybrid IV is unstable with a melting temperature of 20.2° C. Neverthless, each member of this pair forms a stable hybrid with its natural complement, in Hybrids II and III.

PAGE analysis also showed that the two members of the matched pair of SBC 28-mers do not hybridize in a stable manner, and that each SBC ODN and its natural complement form a stable hybrid. Moreover, the normal Watson strand showed no preference for the normal Crick strand over the SBC Crick strand because when equimolar amounts of these three strands were mixed simultaneously at room temperature about equal amounts of the duplex Hybrids I and III were formed. Additionally, there was little, if any, strand displacement or strand exchange when pre-formed Hybrid III was incubated with the normal homolog of the SBC strand, or with the Hybrid II. These data demonstrate that the SBC ODNs behave like natural ODNs when hybridized with as SBC(V) and SBC(VI) are modified so that each dA and each dT is replaced with the d2amA and d2sT, respectively. The melting temperatures of these pairs are indicated in the Table.

TABLE 2

Sequence ID NO: 5	
ODN V	5'-GTAAGAGAATTATGCAGTGC-3'
Sequence ID NO: 6	
ODN VI	3'-CATTCTCTTAATACGTCACG-5'
Sequence ID NO: 7	
SBC(V)	5'-G2sT2amA2amAG2amAG2amA2amA2sT2sT2amA2sTGC2amAG2sTGC-3'
Sequence ID NO: 8	
SBC(VI)	3'-C2amA2sT2sTC2sTC2sT2sT2amA2amA2sT2amACG2sTC2amACG-5'

	MELTING TEMPERATURE OF HYBRIDS			
	ODN(V)	ODN(VI)	SBC(V)	SBC(VI)
ODN(V)	—	55° C.	—	64° C.
ODN(VI)	55° C.	—	65° C.	—
SBC(V)	—	65° C.	—	26° C.
SBC(VI)	64° C.	—	26° C.	—

As it can be seen, the ODNs fully modified with the preferred A' and T' modifications of the present invention exhibit even stronger binding to the natural complementary ODNs than the binding between two natural complementary strands. At the same time, the matched pair of SBC ODNs are nevertheless incapable of forming a stable hybrid with each other (their melting temperature is 26° C.

Additional experiments conducted in accordance with the present invention, in terms of melting temperature measurements and PAGE analysis, showed that a matched pair of SBC ODNs complementary to both strands of a target sequence of double stranded DNA is capable of invading the natural duplex nucleic acid to give a stable 3-armed joint. Analogous paired normal DNA ODNs failed to invade the same target. In case of long double stranded DNA, sequentially hybridizing the paired SBC ODNs to each member of

was kept at 80–90° C. for 5 hours and then evaporated to a syrup. The syrup was dissolved in dichloromthane (10 mL) and added to ice cold methanolic ammonia (100 mL) in a glass pressure bottle. After two days at RT the contents of the bottle were evaporated to dryness. The residue was dissolved in methanol and adjusted to pH 8 with freshly prepared sodium methoxide to complete the deprotection. After stirring overnight the solution was treated with Dowex^R-50 H+ resin, filtered and evaporated to dryness. The residue was chromatographed on silica gel using acetone/hexane (3/2) as eluent to give 2.0 g (77%) of analytically pure product.

Example 7

1-(2-Deoxy-β-D-erythropentofuranosyl)-3-[5-(tritylamino)-pentyl]pyrazolo[3,4-d]pyrimidin-4-amine 5'-monophosphate.

To an ice cold solution of the pyrazolopyrimidin-4-amine of Example 6 (250 mg, 0.43 mmole) in trimethyl phosphate (5 mL) was added phosphoryl chloride (50 μL) and the solution was kept at 0–4° C. The reaction was monitored by reversed phase HPLC using a linear gradient from 0 to 100% acetonitrile in water over 25 min. After stirring for 5 hours, an additional aliquot of phosphoryl chloride (25 μL) was added and the solution was stirred another 30 min. The solution was poured into 0.1M ammonium bicarbonate and kept in the cold overnight. The solution was then extracted with ether and the aqueous layer evaporated to dryness. The residue was dissolved in water (5 mL) and purified by reversed phase HPLC using a 22 mm×50 cm C18 column. The column was equilibrated in water and eluted with a gradient of 0 to 100% acetonitrile over 20 min. Fractions containing the desired material were pooled and lyophilized to give 160 mg (56%) of chromatographically pure nucleotide.

Example 8

1-(2-Deoxy-β-D-erythropentofuranosyl)-3-{5-[(6-biotinamido)hexanamido]pentyl}pyrazolo[3,4-d]pyrimidin-4-amine 5'-monophosphate.

An ethanol solution (10 mL) of the nucleotide of Example 7, palladium hydroxide on carbon (50 mg), and cyclohexadiene (1 mL) was refluxed for 3 days, filtered, and evaporated to dryness. The residue was washed with dichloromethane, dissolved in DMF (1.5 mL) containing triethylamine (100 mL), and treated with N-hydroxysuccinimidyl biotinylaminocaproate (50 mg). After stirring overnight an additional amount of N-hydroxysuccinimidyl 6-biotinamidocaproate (50 mg) was added and the solution was stirred for 18 hours. The reaction mixture was evaporated to dryness and chromatographed following the procedure in Example 7. Fractions were pooled and lyophilized to give 80 mg of chromatographically pure biotinamido-substituted nucleotide.

Example 9

1-(2-Deoxy-β-D-erythropentofuranosyl)-3-[5-(6-biotinamido)-hexanamidopentyl]pyrazolo[3,4-d]pyrimidin-4-amine 5'-triphosphate.

The monophosphate of Example 8 (80 mg, ca. 0.1 mmole) was dissolved in DMF with the addition of triethylamine (14 μL). Carbonyldiimidazole (81 mg, 0.5 mmole) was added and the solution stirred at RT for 18 hours. The solution was treated with methanol (40 μL), and after stirring for 30 minutes tributylammonium pyrophosphate (0.5 g in 0.5 mL DMF) was added. After stirring for 24 hours another aliquot of tributylammonium pyrophosphate was added and the solution was stirred overnight. The reaction mixture was evaporated to dryness and chromatographed following the procedure in Example 8. Two products were collected and were each separately treated with conc. ammonium hydroxide (1 mL) for 18 hours at 55° C. UV and HPLC analysis indicated that both products were identical after ammonia treatment and were pooled and lyophilized to give 35.2 mg of nucleoside triphosphate.

Example 10

Nick-Translation Reaction

The triphosphate of Example 9 was incorporated into pHPV-16 using the nick translation protocol of Langer et al. (supra). The probe prepared with the triphosphate of Example 9 was compared with probe prepared using commercially available bio-11-dUTP (Sigma Chemical Co). No significant differences could be observed in both a filter hybridization and in in situ smears.

More specifically, the procedure involved the following materials and steps

Materials:

DNase (ICN Biomedicals)—4 μg/mL

DNA polymerase 1 (U.S. Biochemicals)—8 U/mL

pHPV-16—2.16 mg/mL which is a plasmid containing the genomic sequence of human papillomavirus type 16.

10x-DP—1M Tris,pH 7.5(20 mL); 0.5M OTT(80 mL); 1M $MgCl_2$(2.8 mL); H_2O (17mL)

Nucleotides—Mix A—2 mM each dGTP, dCTP, TTP (Pharmacia) Mix U—2 mM each dGTP, DcTP, dATP

Bio-11-dUTP—1.0 mg/mL (BRL)

Bio-12-dAPPTP—1.0 mg/mL

Steps:

To an ice cold mixture of 10x-DP (4 mL), pHV-16 (2 mL), nucleotide mix A (6 mL), Bio-12-dAPPTP (2 mL), and H_2O (20 mL) was added DNase (1 mL) and DNA polymerase 1 (2.4 mL). The reaction mixture was incubated at 16° C. for 1 hour. The procedure was repeated using Bio-11-dUTP and nucleotide mix U in place of Bio-12-dAPPTP (comprising the triphosphate of Example 9) and nucleotide mix A.

Nucleic acid was isolated by ethanol precipitation and hybridized to pHPV-16 slotted onto nitrocellulose. The hybridized biotinylated probe was visualized by a streptavidin-alkaline phosphatase conjugate with BCIP/NBT substrate. Probe prepared using either biotinylated nucleotide gave identical signals. The probes were also tested in an in situ format on cervial smears and showed no qualitative differences in signal and background.

Example 11

5-Amino-3-[(5-tritylamino)pentyl]pyrazole-4-carboxamide.

Following the procedure of Example 2, except that cyanoacetamide is used instead of malononitrile, 5-(tritylamino)pentylhydroxymethylececyanoacetamide is prepared from 6-(tritylamino)caproic acid. This is then treated with diazomethane to give the methoxy derivative, following the procedures of Example 3, which is then reacted with hydrazine monohydrate, as in Example 4, to give 5-amino-3-[(5-tritylamino)-pentyl]pyrazole-4-carboxamide.

Example 12

4-Hydroxy-6-methylthio-3-[(5-tritylamino)pentyl]pyrazolo-[3,4-d]pyrimidine.

The carboxamide from Example 11 is reacted with potassium ethyl xanthate and ethanol at an elevated temperature to give the potassium salt of 4-hydroxypyrazolo[3,4-d] pyrimidine-6-thiol. This salt is then reacted with

oligonucleotides came off at 10 minutes; amino derivatives took 11–12 minutes. The desired oligonucleotide was collected and evaporated to dryness, then it was redissolved in 80% aqueous acetic acid for 90 minutes to remove the trityl group. Desalting was accomplished with a G25 Sephadex column and appropriate fractions were taken. The fractions were concentrated, brought to a specific volume, dilution reading taken to ascertain overall yield and an analytical HPLC done to assure purity. oligonucleotides were frozen at −20° C. until use.

In general, to add the crosslinking arm to an aminoalkyloligonucleotide, a solution of 10 μg of the aminoalkyloligonucleotide and a 100X molar excess of n-hydroxysuccinimide haloacylate such as α-haloacetate or 4-halobutyrate in 10 μL of 0.1M borate buffer, pH 8.5, is incubated at ambient temperature for 30 min. in the dark. The entire reaction is passed over a NAP-10 column equilibrated with and eluted with distilled water. Appropriate fractions based on UV absorbance are combined and the concentration is determined spectrophotometrically.

2,3 5,6-Tetrafluorophenyl trifluoroacetate.

A mixture of 2,3,5,6-tetrafluorophenol (55.2 g, 0.33 mol), trifluoroacetic anhydride (60 mL, 0.42 mol) and boron trifluoride etherate (0.5 mL) was refluxed for 16 hr. Trifluoroacetic anhydride and trifluoroacetic acid were removed by distillation at atmospheric pressure. The trifluoroacetic anhydride fraction (bp 40° C.) was returned to the reaction mixture along with 0.5 mL of boron trifluoride etherate, and the mixture was refluxed for 24 hr. This process was repeated two times to ensure complete reaction. After distillation at atmospheric pressure, the desired product was collected at 62° C./45 mm (45° C./18 mm) as a colorless liquid: yield=81.3 g (93%); d=1.52 g/mL; n_D^{21}=1.3747; IR ($CHCl_3$) 3010, 1815, 1525, 1485, 1235, 1180, 1110, and 955 cm^{-1}. Anal. Calcd for $C_8HF_7O_2$: C, 36.66; H, 0.38; F, 50.74. Found: C, 36.31; H, 0.43; F, 50.95.

2,3,5 6-Tetrafluorophenyl-4'-[bis(2-chloroethyl)amino] phenylbutyrate (Chlorambucil 2,3,5,6-tetrafluorophenyl ester)

To a solution of 0.25 g (0.82 mmol) of chlorambucil (supplied by Fluka A. G.) and 0.3 g (1.1 mmol) of 2,3,5,6-tetrafluorophenyl trifluoroacetate in 5 ml of dry dichloromethane was added 0.2 Ml of dry triethylamine. The mixture was stirred under argon at room temperature for 0.5 h and evaporated. The residual oil was purified by column chromatography on silica gel with hexane-chloroform (2:1) as the eluting solvent to give the ester as an oil: 0.28 g (75%); TLC on silica gel ($CHCl_3$) R_f0.6; IR (in $CHCl_3$) 3010, 1780, 1613, 1521, 1485 cm^{-1}.

2-Propargyloxyethyl)amine (John, R., and Seitz, G., Chem. Ber., 123, 133 (1990) was prepared by condensing propynol with 2-bromoethylammonium bromide in liquid ammonia in the presence of Na NH_2, and was used crude for the next reaction.

3-(2-Trifluoroacetamidoethoxy)propyne

(2-Propargyloxyethyl)amine (13.8 g, 0.14 mol) is stirred and chilled in an iso-propanol-dry ice bath while excess of trifluoroacetic anhydride (26 ml, 0.18 mol) is added dropwise. N-(2-Propargyloxyethyl)trifluoroacetamide is distilled at 84–85°/1.7 torr as an oil which solidified upon refrigeration; yield 14.4 g (52%), m.p. (160, n_D^{24} 1.4110. Anal. Calcd. for $C_7H_8F_3NO_2$: C, 43.09, H, 4.13; N, 7.18; F, 29.21. Found: C, 42.80; H, 4.03; N, 7.06; F, 29.38.

5-[3-(2-Trifluoroacetamidoethoxy),Propynyl]-2'-deoxyuridine

A mixture of 5-iodo-2'-deoxyuridine (3.54 g, 10 mmol), copper(I) iodide (0.19 g, 1 mmol) and tetrakis (triphenylphosphine)palladium(O) (0.58 g, 0.5 mmol) is dried in vacuo at 60° for 3 hours and placed under argon. A suspension of the mixture in dry DMF (20 ml) is stirred under argon and treated with dry triethylamine (1.7 ml, 12 mmol) followed by 3-(2-Trifluoroacetamidoethoxy)propyne (3.17 g, 16 mmol). The mixture is cooled at room temperature in a water bath and stirred for 17 hours. The mixture is treated with 2% acetic acid (100 ml), the catalyst is removed by filtration and washed with 50% methanol. The filtrates are combined and passed onto a LiChroprep RP-18 column (5x25 cm), the column is washed, then eluted with 1% acetic acid in 50% (v/v) methanol. The fractions with the main product are combined, evaporated, and dried in vacuo. The resultant foam is stirred with 150 ml of ether to give crystalline product; yield 3.6 g (85%); m.p. 145–1520.

5-[3-(2-Trifluoroacetamidoethoxy)propyl]2'-deoxyuridine

A solution of 5-[3-(2-trifluoroacetamidoethoxy) propynyl]-2'-deoxyuridine (3.4 g, 8.1 mmol) in methanol (20 ml) is stirred with ammonium formate (prepared by addition of 3 ml, 79 mmol of cold 98% formic acid into 2 ml, 50 mmol of dry ice frozen 25% ammonia) and 0.2 g of 10% Pd/C for 7 hours at room temperature under hydrogen atmosphere. The catalyst is removed by filtration, the filtrate evaporated and product is purified on LiChroprep RP-18 column by the above procedure. Fractions containing the desired product are combined and evaporated to dryness in vacuo and the resultant solid is triturated with dry ether to give 3.0 g (87% product, m.p. 107–110°; max in nm, in 0.1M triethylamine-acetate (pH 7.5), 220, 268. Analysis calculated for $C_{16}H_{22}F_3N_3O_7$: C, 45.18; H, 5.21; N, 9.88; F, 13.40. Found C, 45.16; H, 5.16; N, 9.68; F, 13.13.

Introduction of chlorambucil residue into the primary amino groups of oligonucleotides

Preparation of the cetyltrimethylammonium salt of oligonucleotides: a 100 μL aliquot of aqueous solution of oligonucleotide (50–500 ug), generally triethylammonium salt, was injected to a column packed with Dowex 50wx8 in the cetyltrimethylammonium form and prewashed with 50% alcohol in water. The column was eluted by 50% aqueous ethanol (0.1 mL/min). Oligonucleotide containing fraction was dried on a Speedvac over 2 hours and used in following reactions.

Ethanol solution (50 uL) of cetyltrimethylammonium salt of an oligonucleotide (50–100 μg) was mixed with 0.08M solution of 2,3,5,6-tetrafluorophenyl-4'-[bis(2-chloroethyl) amino]phenylbutyrate (tetrafluorophenyl ester of chlorambucil) in acetonitrile (50 μL) and 3 μL of diisopropylethylamine. After shaking for three hours at room temperature, the product was precipitated by 2% $LiClO_4$ in acetone (1.5 mL). The product was reprecipitated from water (60 uL) by 2% $LiClO_4$ in acetone three times. Finally the chlorambucil derivative of the oligonucleotide was purified by Reverse Phase Chromatography with approximately 50–80% yield. The fraction containing the product was concentrated by addition of butanol. The isolated chlorambucil derivative of the oligonucleotide was precipitated in acetone solution with $LiClO_4$, washed by acetone and dried under vacuum. All manipulations of reactive oligonucleotide were performed as quickly as possible, with the product in ice-cold solution.

Preparation of SBC ODNs

N-phenoxyacetyl protected 2'-deoxyguanosine and 2'-deoxycytidine 3'-O-2-cyanoethyl-N,N'-diisopropylphosphoramidite are available commercially from BioGenex, Alameda, California. 5'-O-dimethoxytrityl-2-thiothymidine-3'-O-(2-cyanoethyl-N,N'-diisopropylphosphoramidite) was prepared using the proce-

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 8

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..28
(D) OTHER INFORMATION: /note= "corresponds to "Watson" strand of Hybrids I & III"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CTGACAACGA TCGGAGGACC GAAGGAGC 28

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..28
(D) OTHER INFORMATION: /note= "corresponds to "Crick" strand of Hybrids I & II"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GCTCCTTCGG TCCTCCGATC GTTGTCAG 28

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: one-of(1, 5, 8, 12, 19, 20, 28)
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "pyrrolo-[2,3-d]pyrimidine-2(3H)-one"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: one-of(3, 9, 13, 14, 16, 17, 21, 24, 25, 27)
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "hypoxanthine"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 1..28
(D) OTHER INFORMATION: /note= "corresponds to "Watson" strand of Hybrids II & IV"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

NTNANAANNA TNNNANNANN NAANNANN 28

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

-continued

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: one-of(3, 8, 10, 11, 19)
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "d2amAdenine replaces all dAdenine"

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: one-of(5, 9, 12, 13, 15, 17, 18)
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "d2sThymine replaces all dThymine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCACTGCATA ATTCTCTTAC 20

What is claimed is:

1. A pair of oligonucleotides (ODNs), each of said ODNs comprising nucleotide moieties having naturally occurring aglycon bases and a combination of modified aglycon bases selected from the group consisting of the combinations (1) A', T', (2) G', C' and (3) A', T', G', C', the duplex form of said pair of ODNs having a melting temperature under physiological conditions of less than approximately 40° C., each of said pair of ODNs being substantially complementary in the Watson-Crick sense to one of the two strands of a duplexed target sequence in nucleic acid,

wherein the nucleotide moieties having the modified bases have the following properties:

within complementary oligonucleotides A' does not form a stable hydrogen bonded base pair with T' and forms a stable hydrogen bonded base pair with T;

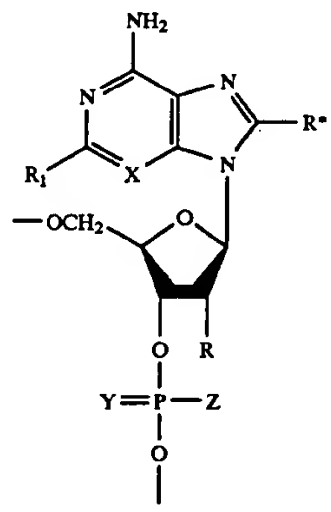
within complementary oligonucleotides T' does not form a stable hydrogen bonded base pair with A' and forms a stable hydrogen bonded base pair with A;

within complementary oligonucleotides G' does not form a stable hydrogen bonded base pair with C' and forms a stable hydrogen bonded base pair with C, and

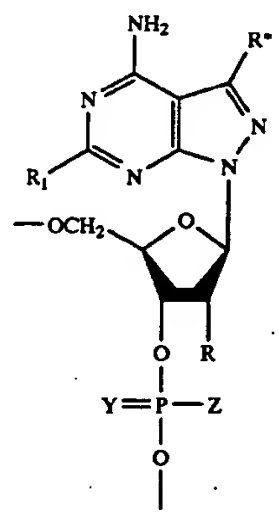
within complementary oligonucleotides C' does not form a stable hydrogen bonded base pair with GI and forms a stable hydrogen bonded base pair with G.

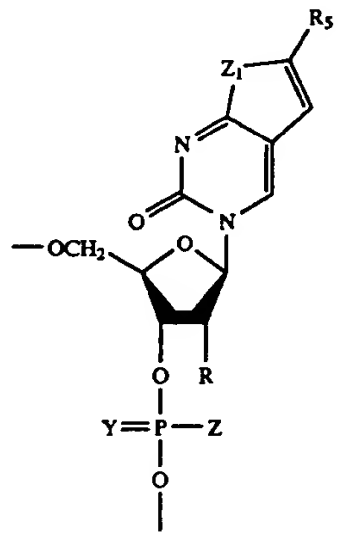
2. The ODNs of claim 1 wherein the nucleotide moiety A' has the structure selected from the groups shown by formulas (i), (ii) and (iii)

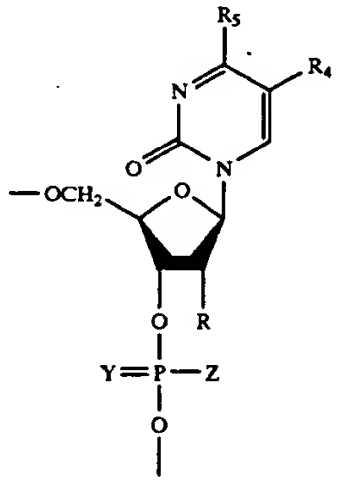
(i)



(ii)







wherein

Y is O or S;

Z is OH or CH_3;

R is H, F, or OR_2, where R_2 is H, C_{1-6} alkyl or allyl,

R_4 is H, C_{1-6} alkyl, C_{1-6} alkenyl, C_{1-6} alknyl, a cross-linking function or a reporter group;

(viii)

Z_1 is O or NH, and

R_5 is H, or C_{1-4} alkyl.

6. The ODNs of claim 2 wherein the nucleotide moiety A' has the structure in accordance with formula (i).

7. The ODNs of claim 6 wherein X is N, Z is OH, and Y is O.

8. The ODNs of claim 7 wherein R_1 is NH_2.

9. The ODNs of claim 3 wherein Z is OH, and Y is O.

10. The ODNs of claim 9 wherein R_4 is CH_3.

11. The ODNs of claim 4 wherein the nucleotide moiety G' has the structure in accordance with formula (v).

12. The ODNs of claim 11 wherein X is N, Z is OH, and Y is O.

13. The ODNs of claim 12 wherein R_1 is H.

14. The ODNs of claim 5 wherein the nucleotide moiety C' has the structure in accordance with formula (viii).

15. The ODNs of claim 14 wherein, Z is OH, Z_1 is NH and Y is O.

16. The ODNs of claim 15 wherein R_5 is H.

(ix)

17. The ODNs of claim 1 having approximately 5 to 99 nucleotide units.

18. The ODNs of claim 1 wherein each of the nucleotides is a 2'-deoxyribonucleotide.

19. The ODNs of claim 1 wherein each of the nucleotides is a ribonucleotide.

20. The ODNs of claim 1 comprising at least one nucleotide unit having a 2-O-methylribose moiety.

21. The ODNs of claim 1 comprising a cross-linking function covalently attached to at least one nucleotide unit.

22. The ODNs of claim 1 comprising a reporter group.

23. The ODNs of claim 1 wherein the combination of modified aglycon bases is A', T'.

24. The ODNs of claim 1 wherein the combination of modified aglycon bases is G', C'.

25. The ODNs of claim 1 wherein the combination of modified aglycon bases is A', T', G', C'.

26. The ODNs of claim 1 where the pair of oligonucleotides are linked to one another by a covalently bonded tether.

* * * * *

NUCLEOSIDES, NUCLEOTIDES & NUCLEIC ACIDS, 20(4–7), 501–506 (2001)

ANALOGS OF GUANINE NUCLEOSIDE TRIPHOSPHATES FOR SEQUENCING APPLICATIONS

Mark G. McDougall, Lei Sun, Inna Livshin, Louis P. Hosta, Bernard F. McArdle, Sui-Bi Samols, Carl W. Fuller, and Shiv Kumar*

Amersham Pharmacia Biotech, 800 Centennial Ave., Piscataway, New Jersey 08855

ABSTRACT

We have synthesized more than 30 different deoxyribonucleosides and triphosphates with modifications either in the base or the phosphate moiety as analogs of 2′-dGTP for DNA sequencing applications. All the modified nucleoside triphosphates were tested as substrates for DNA polymerases, including Sequenase™ T7 DNA polymerase or Thermo Sequenase™ DNA polymerase. Two of the analogs, 7-ethyl-7-deaza-dGTP and 7-hydroxymethyl-7-deaza-dGTP meet our requirements as better sequencing reagents.

One of the most common difficulties in DNA sequencing is known as the "Compression" artifact. This apparently results when sequence-dependent secondary structures form in the DNA as it is being analyzed by gel electrophoresis (1). Fragments with these secondary structures migrate faster than similar-size fragments, which lack the secondary structure. There are several techniques used to eliminate these artifacts including electrophoresis at elevated temperature, the addition of higher concentration of urea and formamide to gels and use of nucleotide analogs. In particular, the dGTP analogs 2′-deoxyinosine-5′-triphosphate (dITP, **22**) and 7-deaza-2′-deoxyguanosine-5′-triphosphate (2) (dZTP, **8**) are commonly used

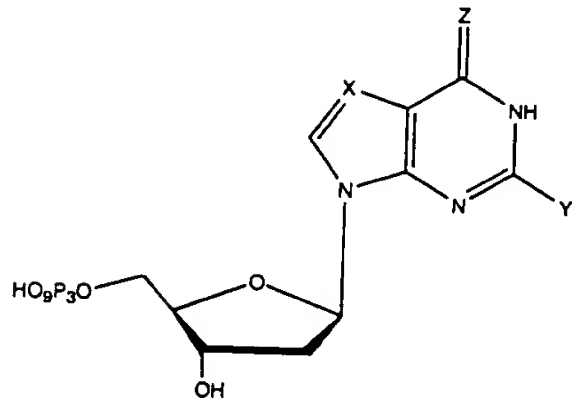
*Corresponding author. E-mail: shiv.kumar@am.apbiotech.com

for sequencing. Both these analogs have specific problems, which prompted us to undertake this line of research.

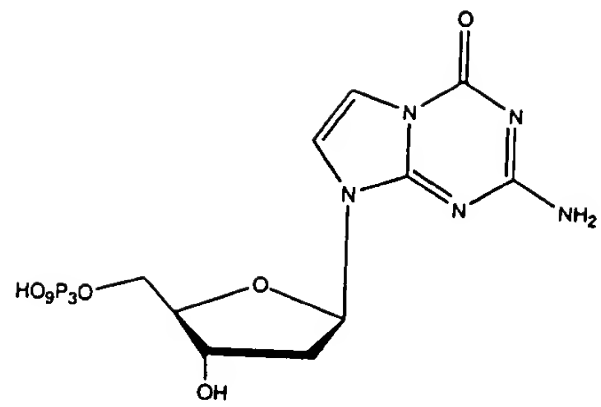
Although dITP has excellent compression resolving properties, eliminating essentially all compressions, the reactivity with both Sequenase™ T7 DNA polymerase and Thermo Sequenase™ DNA polymerase is low (10 to 20% that of dGTP), and uniformity of terminations in the sequencing reactions exhibit sequence specific variations, which can be severe. This makes sequence interpretation more difficult and certain sequencing applications impossible. 7-Deaza-2′-deoxyguanosine-5′-triphosphate has good reactivity with DNA polymerases and reasonable uniformity. However, it resolves only 75–80% of all compressions in slab gels and less in capillary electrophoresis where the denaturing conditions are weaker. Therefore, the goal of this research was to find an analog with better compression resolution than dZTP and better reactivity and uniformity than dITP.

In a research program directed towards finding better substrates for the DNA polymerase than dITP and better compression resolution properties than 7-deaza-dGTP, we have prepared a number of 2′-deoxynucleoside-5′-triphosphates as novel analogs of the guanine bases (Fig. 1). Most of the nucleosides were prepared by using literature procedures with minor modifications and the triphosphates were prepared using $P(O)Cl_3$ and bis(tri-*n*-butylammonium)pyrophosphate in a "one-pot, three-step" phosphorylation procedure (3). The compounds 6-thio dITP (**2**) and 6-dGTP (**3**) were prepared as previously reported (4). The nucleosides 2-bromo- and 2-fluoro-2′-deoxyinosine were synthesized according to the methods developed by Robins (5). 5-Aza-7-deaza-2′-deoxyguanosine (6) and 3,7-dideaza-2′-deoxyguanosine (7) were synthesized using minor modifications of published procedures. 7-Halogenated-7-deaza dGTP analogs have been previously reported (8). The alkynyl series of 7-deaza-2′-deoxyguanosine was prepared according to the literature methods (9,10). The synthesis of 7-alkyl- and 7-hydroxyalkyl-7-deaza-2′ deoxyguanosine as well as 7-substituted-7-deaza-2′-deoxyinosine nucleosides will be reported elsewhere (11). All these nucleoside triphosphates were tested for activity as substrates for DNA polymerase. The analog triphosphates that proved active were also tested for sequencing on templates with known compression artifacts and scored relative to dGTP, dZTP and dITP. Specifically, these analogs were used as substrates replacing dGTP in dye primer cycle sequencing reactions with Thermo Sequenase™ on double stranded pGEM 3Zf(+), starting at the −28 reverse primer. The ratio of an analog to ddGTP was varied in the sequencing reaction mixtures, which were the same as typically used in standard primer labeled sequencing reactions for dGTP or dZTP (12). The results of these experiments are given in Table 1.

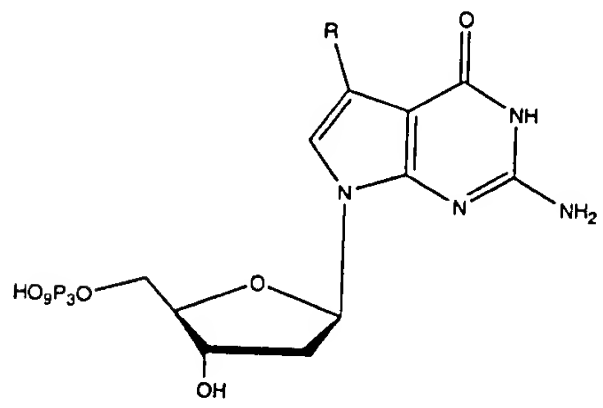
Three characteristics for each analog triphosphate are recorded in Table 1. Relative reactivity is a rough assessment of the compound as substrates for the exonuclease free Thermo Sequenase™ DNA polymerase. These values were determined from the optimal sequencing reaction ratios, reflecting competition between the analog and ddGTP. The comment word "stops" indicate the inability of the polymerase to extend past sites of multiple analog incorporation. Band uniformity is a term used for changes in the sequence specific incorporation of ddGTP, as



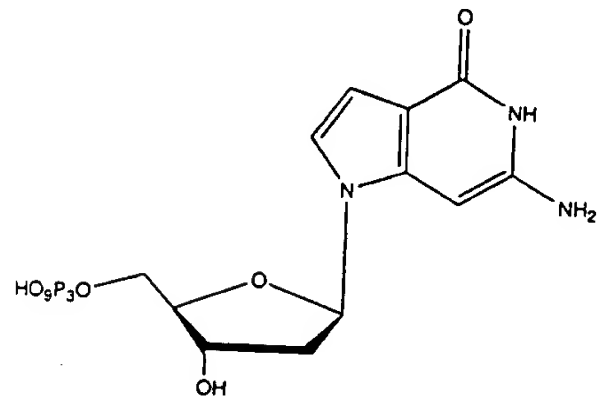
dGTP (1) X = N, Y = NH_2, Z = O
6-thio-dITP (2) X = N, Y = H, Z = S
6-thio-dGTP (3) X = N, Y = NH_2, Z = S
2-bromo-dITP (4) X = N, Y = Br, Z = O
2-fluoro-dITP (5) X = N, Y = F, Z = O
7-deaza-dGTP (8) X = CH, Y = NH_2, Z = O
dITP (22) X = N, Y = H, Z = O
7-propynyl-7-deaza-dITP (23) X = C-C≡CCH_3, Y = H, Z = O
7-propyl-7-deaza-dITP (24) X = C-$CH_2CH_2CH_3$, Y = H, Z = O



5-aza-7-deaza dGTP (6)



R = Cl (9), I (10), CH_3 (11), CH_2CH_3 (12), $CH_2CH_2CH_3$ (13), $CH(CH_3)_2$ (14), $(CH_2)_5CH_3$ (15), C≡CCH_3 (16), C≡$C(CH_2)_3CH_3$ (17), CH_2OH (18), CH_2CH_2OH (19), $CH_2CH_2CH_2OH$ (20), C≡CCH_2OH (21).



3,7-dideaza dGTP (7)

Figure 1.

observed by variations in band intensity. The values reported are the variance of band intensities normalized by an algorithm. Absolute uniformity is scored as zero. Compression scores record the analog's ability to relieve compression artifacts, the larger the number the greater the ability to relieve. The scores of zero for dGTP and four for dITP were arbitrarily set from a limited number of compression artifacts to ascertain the trend of this property.

The nucleoside triphosphates with modifications to the hydrogen bonding face (**2–7**) were all unsuitable for sequencing applications. All these compounds, with the exception of 6-thio dGTP (**3**), had very poor to no reactivity with the DNA polymerases tested. Surprisingly, the absence of N1(H) in a G analog, either by design (**6**) or by ionization (**5**) led to inactivity. The nucleotide, 2-F dITP (**5**) was found to have a pKa value of 5.2 as determined by UV absorbance titration and curve fitting. The optimal sequencing conditions for Thermo Sequenase™ DNA polymerase is at pH = 9.5. At this pH, extension with this analog is almost non-existent, however

Table 1. Results of Primer Labeled Sequencing Using Thermo Sequenase™ Polymerase

Compound	Relative Reactivity	Band Uniformity[1]	Compression Score[2]
1. dGTP	1	0.15	0
2. 6S-dITP	0.1(stops)		
3. 6S-dGTP	0.9(stops)	0.66	0
4. 2Br-dITP	0.1(stops)		
5. 2F-dITP	0.1 to 0.2 (at pH = 9.5 to 7.5)		
6. 5aza-7deaza-dGTP	0		
7. 3,7 dideaza-dGTP	0		
8. 7-Deaza-dGTP(dZTP)	1	0.19	1
9. 7-Chloro-dZTP	1	0.23	−1
10. 7-Iodo-dZTP	1.1	0.23	−1
11. 7-Methyl-dZTP	1	0.18	2
12. 7-Ethyl-dZTP	1	0.19	3
13. 7-Propyl-dZTP	0.85	0.21	3.5
14. 7-Isopropyl-dZTP	0.2(stops)		4
15. 7-Hexyl-dZTP	0.1(stops)		4
16. 7-Propynyl-dZTP	1.25	0.21	0
17. 7-Hexynyl-dZTP	0.6(stops)		1
18. 7-Hydroxymethyl-dZTP	1	0.16	1.5
19. 7-Hydroxyethyl-dZTP	0.2(stops)		
20. 7-Hydroxypropyl-dZTP	0.6	0.64	3.5
21. 7-Hydroxypropynyl-dZTP	0.5	0.28	0
22. dITP	0.2	0.33	4
23. 7-Propyl-7-deaza dITP	0.4	0.25	4
24. 7-Propynyl-7-deaza dITP	0.1 (stops)	0.3	4

[1]Lower numbers more uniform heights.
[2]Fewer compressions, higher score.

somewhat better activity is observed upon lowering the pH to 7.5. A related pH phenomenon is observed in the thermodynamic properties of duplexes containing 2-fluoro-2′-deoxyinosine (13).

Compound **7** is an example of the importance of the nitrogen at the N-3 position for polymerase substrate recognition (14). Although, 6-thio dGTP (**3**) exhibits good activity under our sequencing conditions, stops or pauses were observed at multiple runs of G incorporation. Such behavior renders this analog unsuitable as a replacement for dGTP in most sequencing applications (4).

Nucleotide derivatives of 7-deaza-2′-deoxyguanosine triphosphate with substitution at the 7-position (**9–21**) yield an exciting class of polymerase substrates. 7-Halogenated (8) (**9–10**) as well as 7-propynyl dZTP (**16**) are excellent substrates, which easily sequence out to 500 bases with intermediate band uniformity. Using these analogs unfortunately made compression artifacts as bad or worse than dGTP. It is likely that the structural properties that impart duplex stability to double stranded DNA containing these bases (9,15), are also responsible for stabilizing the structures which result in compression artifacts. Increasing the size of the substituent on the triple bond as in compounds **17** and **21** lowered substrate activity.

with 7-hexynyl dZTP (**17**) exhibiting abortive behavior at multiple G incorporation sites.

Alkyl derivatives of 7-deaza-2′-deoxyguanosine triphosphate (**11–15**) display a trend where increasing size gives increasing compression resolution but decreasing polymerase reactivity. 7-Methyl (**11**) and 7-ethyl dZTP (**12**) have a relative reactivity that matches the parent compound (**8**). 7-Propyl dZTP (**13**) shows less reactivity and more background noise in primer sequencing reactions. However, the ability to resolve compression artifacts is in the order: propyl > ethyl > methyl > hydrogen. The increase in steric bulk at the 7-position, such as with isopropyl (**14**) or chain length as with hexyl (**15**) diminishes the reactivity greatly and elongation at specific sites is terminated.

The hydroxyalkyl series (**18–20**) has very interesting polymerase behavior. 7-Hydroxymethyl-7-deaza-2′-deoxyguanosine-5′-triphosphate (**18**) with reactivity and uniformity as good as 2′-dGTP (**1**) and a better compression score than dZTP (**8**) was by far the best sequencing analog tested. Under difficult template conditions, this analog out-performed any other analog tested in this study. Extending the alkyl chain by one methylene group, to hydroxyethyl (**19**), led to poor reactivity and polymerase stops. Hydroxypropyl dZTP (**20**) did not terminate polymerase extentions but has large uniformity problems. We believe that the hydrogen bonding capability in these analogs is responsible for their polymerase behavior and more studies are underway to understand this class of molecules.

Two compounds (**23–24**) were prepared as analogs of 7-substituted-7-deaza-dITP. Initial results indicate that the 7-propynyl derivative (**23**) is as good or better a substrate as 2′-dITP for Thermo Sequenase™ DNA polymerase. The propynyl substitution also helps band uniformity while retaining an excellent compression score. Further testing of this analog will be done to evaluate its utility for routine sequencing with both dyelabeled primers and dye-labeled terminators in slab gels and capillary electrophoresis. The alkyl addition to 7-deaza-7-dITP, namely 7-Propyl-7-deaza-dITP (**24**) worsened reactivity of an already poor substrate.

In conclusion, we have synthesized a number of analogs of 2′-deoxyguanosine triphosphate. Our results show that 7-ethyl, 7-propyl, and 7-hydroxymethyl-7-deaza-2′-deoxyguanosine triphosphates are good replacements for 2′-dGTP in sequencing reactions, eliminating compression artifacts to a greater degree than the parent compound, 7-deaza-2′-deoxyguanosine triphosphate. We are now in the process of determining why certain analogs relieve compression artifacts better than others do and examining other biological activities for our novel nucleosides and nucleotides.

REFERENCES

1. a) Yamakawa, H.; Nakajima, D.; Ohara, O. *DNA Res.*, **1996**, *3*, 81–86. b) Yamakawa, H.; Ohara, O.*Nucl. Acids Res.*, **1997**, *25*, 1311–1312.
2. Mizusawa, S.; Nishimura, S.; Seela, F. *Nucleic Acids Res.*, **1986**, *14*, 1319–1324.

3. Burgess, K.; Cook, D. *Chem. Rev.*, **2000**, *100*, 2047–2059 and references therein.
4. McDougall, M.G.; McArdle, B.A.; Kumar, S.; Fuller, C.W.; Rellick, L.M.; Govorine, A.V. *Nucleosides & Nucleotides* **1997**, *16*, 1745–1748.
5. a) Robins, M.J.; Uznanski, B. *Can. J. Chem.*, **1981**, *59*, 2608–2611. b) Woo, J.; Sigurdsson, S. Th.; Hopkins, P.B. *J. Am. Chem. Soc.*, **1993**, *115*, 3407–3415.
6. a) Kim, S.; Bartholomew, D.; Allen, L. *J. Med. Chem.*, **1978**, 21, 883. b) Rosemeyer, H.; Seela, F. *J. Org. Chem.*, **1987**, *52*, 36–5143.
7. a) Girgis, N.; Cottam, H.; Larson, S. and Robins, R. *Nucleic Acid Research*, **1987**, *15*, 1217–1226. b) Schneller, S.; Hosmane, R.; MacCartney, L. and Hessinger, D., *J. Med. Chem.*, **1978**, *21*, 991.
8. McDougall, M.G.; Hosta, L.; Kumar, S.; Fuller C.W. *Nucleosides & Nucleotides*, **1999**, *18*, 1009–1011.
9. Buhr, C. A.; Wagner, R. A.; Grant, D.; Froehler, B. C. *Nucleic Acids Res.* **1996**, *24*, 2974–80.
10. Ramzaeva, N.; Seela, F. *Helv. Chim. Acta*, **1995**, *78*, 1083–1090.
11. Manuscript in preparation.
12. a) For sequencing conditions using Sequenase™ see: Fuller, C. in *Methods in Enzymology*, Vol. 216, Academic Press: New York, 1992, pp. 329–354. b) For sequencing conditions using Thermo Sequences™ see: Vander Horn, P.B.; Davis, M.C.; Cunniff, J.J.; Ruan, C.; McArdle, B.F.; Samols, S.B.; Szaz, J.; Hu, G.; Hujer, K.M.; Domke, S.T.; Brummet, S.R.; Moffett, R.B.; Fuller, C.W. *BioTechniques*, **1997**, *22*, 758–765.
13. Avino, A.; Garcia, G.G.; Marquez, V.E.; Eritja, R. *Bioorg. Med. Chem. Lett.*, **1995**, *5*, 2331–2336.
14. a) Cosstick, R.; Li, X.; Tuli, D.K.; Williams, D.M.; Connolly, B.A.; Newman, P.C. *Nucleic Acids Res.*, **1990**, *18*, 4771–4778. b) Spratt, T.E. *Biochemistry*, **1997**, *36*, 13292–13297. c) Morales, J.C.; Kool, E.T. *J. Am. Chem. Soc.*, **1999**, *121*, 2323–2324.
15. a) Seela, F.; Ramzaeva, N.; Chen, Y. *Bioorg. Med. Chem. Lett.*, **1995**, *5*, 3049–52. b) Ramzaeva, N.; Seela, F. *Helv. Chim. Acta*, **1996**, *79*, 1549–58.

Nucleic Acids Research, 1998, Vol. 26, No. 21 4975–4982

Enhanced high density oligonucleotide array-based sequence analysis using modified nucleoside triphosphates

Joseph G. Hacia, Stephen A. Woski[1], Jacqueline Fidanza[2], Keith Edgemon, Nathaniel Hunt, Glenn McGall[2], Stephen P. A. Fodor[2] and Francis S. Collins*

National Human Genome Research Institute, National Institutes of Health, Building 31, Room 4B09, 31 Center Drive, Bethesda, MD 20892-2152, USA, [1]Department of Chemistry, University of Alabama, Tuscaloosa, AL 35487, USA and [2]Affymetrix, Santa Clara, CA 95051, USA

Received May 29, 1998; Revised and Accepted September 10, 1998

ABSTRACT

Pairs of high density oligonucleotide arrays (DNA chips) consisting of >96 000 oligonucleotides were designed to screen the entire 5.53 kb coding region of the hereditary breast and ovarian cancer *BRCA1* gene for all possible sequence changes in the homozygous and heterozygous states. Single-stranded RNA targets were generated by PCR amplification of individual *BRCA1* exons using primers containing T3 and T7 RNA polymerase promoter tails followed by *in vitro* transcription and partial fragmentation reactions. Fluorescent hybridization signals from targets containing the four natural bases to >5592 different fully complementary 25mer oligonucleotide probes on the chip varied over two orders of magnitude. To examine the thermodynamic contribution of rU·dA and rA·dT target·probe base pairs to this variability, modified uridine [5-methyluridine and 5-(1-propynyl)-uridine)] and modified adenosine (2,6-diaminopurine riboside) 5′-triphosphates were incorporated into *BRCA1* targets. Hybridization specificity was assessed based upon hybridization signals from >33 200 probes containing centrally localized single base pair mismatches relative to target sequence. Targets containing 5-methyluridine displayed promising localized enhancements in hybridization signal, especially in pyrimidine-rich target tracts, while maintaining single nucleotide mismatch hybridization specificities comparable with those of unmodified targets.

INTRODUCTION

Light-directed combinatorial chemical approaches allow the manufacture of high density arrays consisting of $>10^5$ distinct oligonucleotide species (20 μm feature size) on 1.2×1.2 cm^2 glass surfaces (1–2). Such arrays have been used to screen for mutations and polymorphisms in the *CFTR* gene (3), the HIV-1 reverse transcriptase and protease genes (4), the β-globin gene (5), the mitochondrial genome (6) and the *BRCA1* gene (7). Furthermore, they have been used to monitor gene expression (8), analyze gene function (9), optimize antisense oligonucleotide design (10) and acquire information from orthologous genes in related species (11).

A significant challenge in high density oligonucleotide array-based applications is to develop assay conditions so all fully complementary perfect match oligonucleotide probes of varying sequence content produce robust and specific target hybridization signals. Subsets of perfect match probes could have a greatly diminished hybridization signal due to decreased duplex stability resulting from sequence composition effects and inter- and intramolecular structures in both target and probes. When reliable data from such probes must be generated, hybridization conditions which are suboptimal for the specificity of other probes with robust hybridization signals may have to be employed. Herein, we analyze the affinity and specificity of RNA targets toward >90 000 oligonucleotide probes present on a pair of high density oligonucleotide arrays that scan the entire coding region of the *BRCA1* gene for all possible homozygous and heterozygous sequence changes. To pursue sequence composition effects on target hybridization and to explore possible solutions, we evaluated the effect of incorporating modified nucleoside triphosphates into *BRCA1* RNA target on hybridization signal and single nucleotide mismatch hybridization specificity.

MATERIALS AND METHODS

Synthesis of pyrimidine 5′-triphosphates

The synthesis of 5-(1-propynyl)-uridine was accomplished in three steps from commercially available 5-iodouridine (Sigma). First, treatment with excess acetic anhydride in pyridine produced tri-*O*-acetyl-5-iodouridine. This compound was converted to the 5-propynyl analog using the method of Hobbs (12). The free nucleoside was then generated by reaction with $NaOCH_3$ in methanol followed by desalting over BioRad AG-501 mixed bed resin.

The 5′-triphosphates of 5-(1-propynyl)-uridine and commercially available 5-methyluridine (R.I. Chemical) were synthesized using a two step procedure. First, conversions of the nucleosides

*To whom correspondence should be addressed. Tel: +1 301 496 0844; Fax: +1 301 402 0837; Email: fc23a@nih.gov

Table 1. *BRCA1* exon amplification primers

Exon	T3 Forward Primer Sequence (5′-3′)	T7 Reverse Primer Sequence (5′-3′)
2	d(ATTAACCCTCACTAAAGGGAAGTTGTCATTTTATAAACCTTT)	d(TAATACGACTCACTATAGGGATGTCTTTTCTTCCCTAGTATGT)
3	d(ATTAACCCTCACTAAAGGGAACGAACTTGAGGCCTTATG)	d(TAATACGACTCACTATAGGGATTGGATTTTCGTTCTCACTTA)
5	d(ATTAACCCTCACTAAAGGGACTCTTAAGGGCAGTTGTGAG)	d(TAATACGACTCACTATAGGGATTCCTACTGTGGTTGCTTCC)
6	d(ATTAACCCTCACTAAAGGGACTTATTTTAGTGTCCTTAAAAGG)	d(TAATACGACTCACTATAGGGATTTCATGGACATCACTTGAGTG)
7	d(ATTAACCCTCACTAAAGGGACAACAAAGAGCATACATAGGG)	d(TAATACGACTCACTATAGGGAGGGCTAAGGCAGGAGGACTGCT)
8	d(ATTAACCCTCACTAAAGGGATGTTAGCTGACTGATGATGGT)	d(TAATACGACTCACTATAGGGATCCAGCAATTATTATTAAATAC)
9	d(ATTAACCCTCACTAAAGGGACCACAGTAGATGCTCAGTAAATA)	d(TAATACGACTCACTATAGGGATAGGAAAATACCAGCTTCATAGA)
10	d(ATTAACCCTCACTAAAGGGATGGTCAGCTTTCTGTAATCG)	d(TAATACGACTCACTATAGGGAGTATCTACCCACTCTCTTCTTCAG)
11	d(ATTAACCCTCACTAAAGGGAATTAAATGAAAGAGTATGAGC)	d(TAATACGACTCACTATAGGGAGTGCTCCCAAAAGCATAAA)
12	d(ATTAACCCTCACTAAAGGGAGTCCTGCCAATGAGAAGAAA)	d(TAATACGACTCACTATAGGGATGTCAGCAAACCTAAGAATGT)
13	d(ATTAACCCTCACTAAAGGGAATGGAAAGCTTCTCAAAGTA)	d(TAATACGACTCACTATAGGGATGTTGGAGCTAGGTCCTTAC)
14	d(ATTAACCCTCACTAAAGGGACTAACCTGAATTATCACTATCA)	d(TAATACGACTCACTATAGGGAGTGTATAAATGCCTGTATGCA)
15	d(ATTAACCCTCACTAAAGGGATGGCTGCCCAGGAAGTATG)	d(TAATACGACTCACTATAGGGAACCAGAATATCTTTATGTAGGA)
16	d(ATTAACCCTCACTAAAGGGAATTCTTAACAGAGACCAGAAC)	d(TAATACGACTCACTATAGGGAAAACTCTTTCCAGAATGTTGT)
17	d(ATTAACCCTCACTAAAGGGAGTGTAGAACGTGCAGGATTG)	d(TAATACGACTCACTATAGGGATCGCCTCATGTGGTTTTA)
18	d(ATTAACCCTCACTAAAGGGAGGCTCTTTAGCTTCTTAGGAC)	d(TAATACGACTCACTATAGGGAGACCATTTTCCCAGCATC)
19	d(ATTAACCCTCACTAAAGGGATCTCCGTGAAAAAGAGC)	d(TAATACGACTCACTATAGGGACATTGTTAAGGAAAGTGGTGC)
20	d(ATTAACCCTCACTAAAGGGATATGACGTGTCTGCTCCAC)	d(TAATACGACTCACTATAGGGAGGGAATCCAAATTACACAGC)
21	d(ATTAACCCTCACTAAAGGGAATCATCAGGTGGTGAACAGAAG)	d(TAATACGACTCACTATAGGGAGTGCTGGAACTCTGGGGTTCT)
22	d(ATTAACCCTCACTAAAGGGATCCCATTGAGAGGTCTTGCT)	d(TAATACGACTCACTATAGGGAGAAGACTTCTGAGGCTAC)
23	d(ATTAACCCTCACTAAAGGGACAGAGCAAGACCCTGTCTC)	d(TAATACGACTCACTATAGGGACATTTTAGCCATTCA)
24	d(ATTAACCCTCACTAAAGGGATGATTAGAGCCTAGTCCAGG)	d(TAATACGACTCACTATAGGGATAGCCAGAAGTCCTTTTCAGG)

into the crude 5′-phosphodichlorodates were accomplished using the procedure of Sowa and Ouchi (13). Without isolation, these compounds were directly converted into the triphosphates by the addition of tributylammonium pyrophosphate and tributylamine in DMF (14). The triphosphates were purified by anion-exchange chromatography eluting with a gradient of triethylammonium formate (pH 6.5). Extensive lyophilization and co-evaporation with water provided the desired triphosphates as their triethylammonium salts. Identities of both compounds were confirmed by ^{31}P NMR spectroscopy and negative ion FAB-MS and the purities were determined to be >95% by analytical anion-exchange HPLC.

Data for 5-MeUTP: ^{31}P NMR (202 MHz, D_2O) δ –10.40 (d, J = 19.7 Hz, P_γ), –11.22 (dd, J = 20.4, 2.5 Hz, P_α), –22.84 (unresolved dd, $J_{apparent}$ = 19.8 Hz, P_β); MS (negative ion FAB) *m/z* 497 (100%, $[M^{4-} + 3H^+]$), 479 (24%, $[M^{4-} + 3H^+ - H_2O]$), 331 (82%).

Data for PUTP: ^{31}P NMR (202 MHz, D_2O) δ –9.24 (m, P_α), –10.69 (d, J = 19.3 Hz, P_γ), –22.27 (unresolved dd, $J_{apparent}$ = 19.7 Hz, P_β); MS (negative ion FAB) *m/z* 521 (100%, $[M^{4-} + 3H^+]$).

NMR spectra were obtained on a Brüker AM500 spectrometer. MS were obtained on a VG Autospec mass spectrometer.

Synthesis of diaminopurine 5′-triphosphate (rDTP)

Diaminopurine-5′-monophosphate was synthesized from diaminopurine riboside (Reliable BioPharmaceuticals) using the described procedure (15) and purified on DEAE–Sephadex employing a 0–0.5 M LiCl gradient to give an 80% yield. The title compound (rDTP) was then synthesized from the monophosphate (16) in 80% yield, purified by RP-HPLC in 50 mM TEAA using an acetonitrile gradient and characterized by ^{31}P NMR and MS.

Data for DTP: ^{31}P NMR (D_2O, relative to H_3PO_4) δ –22.4 (t), –10.8 (d), –8.6 (broad, m). LCMS = 521.0 (M-H).

RNA target preparation

In vitro transcription templates were generated by PCR amplification of all *BRCA1* coding exons from genomic DNA using intronic forward and reverse primer pairs containing T3 and T7 promoter sequences, respectively (Table 1). Exon 11 was amplified using the EXPAND™ Long Range PCR Kit (Boehringer Mannheim) (7). The remaining 21 coding exons were amplified using the Amplitaq Gold PCR Kit (Perkin Elmer). Approximately 5 ng of each exon (except exon 11) were pooled and subject to T3 and T7 RNA polymerase *in vitro* transcription reactions. Exon 11 templates (~50 ng) were transcribed in a separate reaction. *In vitro* transcription reactions were performed in 20 µl reaction volumes using T3 RNA polymerase transcription buffer (Promega), 0.7 mM of the appropriate nucleoside triphosphates, 10 mM DTT, 0.7 mM biotin-16-CTP (Enzo Diagnostics) and 10 U T3 or T7 RNA polymerase as indicated. A 10 µl volume of pooled *BRCA1* exons was diluted into a 8 µl solution of 100 mM $MgCl_2$ with exon 11 transcription products separately treated in a like manner. Reactions were incubated at 94°C for 15 and 45 min, respectively, to fragment targets into ~50–100 nt long pieces which are more accessible to hybridization (7). In theory this produces a relatively random distribution of fragmentation products, however, some phosphodiester internucleotide linkages may be more reactive to hydrolysis than others. This may influence target hybridization in some sequence contexts and should be taken into consideration when interpreting array hybridization data in this study.

Array hybridization and data collection

Fragmented exon 11 and pooled exon targets were combined and diluted into a 400 µl volume of hybridization buffer A (3 M tetramethylammonium chloride, 1× TE, pH 7.4, 0.001% Triton X-100) or B (6× SSPE, 0.005% Triton X-100) containing either 1 nM 5′-fluorescein-labeled control oligodeoxyribonucleotides S (5′-CGGTAGCATCTTGAC-3′) or AS (5′-GTCAAGATGC-TACCG-3′) (for arrays complementary to sense and antisense strand targets, respectively) (7). Arrays were hybridized with target, stained with a phycoerythrin–streptavidin conjugate (Molecular Probes) and hybridization signals quantitated as previously described (7).

RESULTS AND DISCUSSION

Oligonucleotide array design

Extending previous analysis of the 3.43 kb central *BRCA1* exon 11 (7), a pair of arrays consisting of >96 600 oligonucleotides was designed to scan both strands of the 5.59 kb *BRCA1* coding sequence (containing 22 coding exons) for all possible sequence

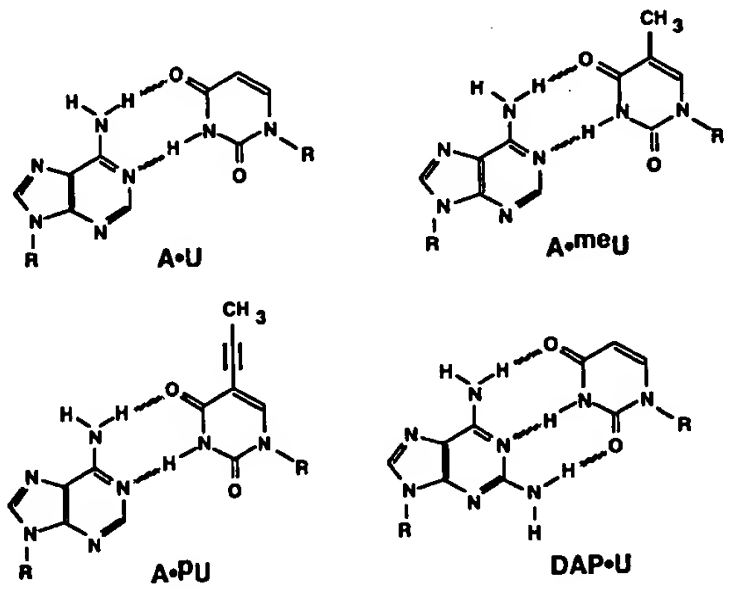


Figure 1. Proposed hydrogen bonding schemes of modified pyrimidine and purine base pairs. ^{me}U, PU and DAP represent 5-methyluridine, 5-(1-propynyl)-uridine and 2,6-diaminopurine, respectively. Dashed lines represent proposed hydrogen bonds.

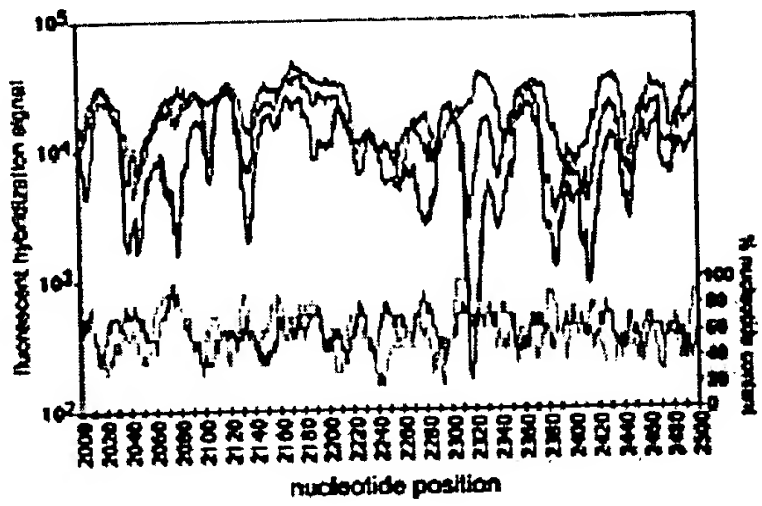


Figure 2. Hybridization signal intensities of modified pyrimidine antisense targets. Averaged fluorescence intensities (two experiments) of targets hybridized (buffer A, 40°C) to arrayed *BRCA1* perfect match probes are shown. Fluorescent signal intensities from perfect match probes complementary to nt 2500–3000 of the *BRCA1* cDNA sequence are plotted on the log scale *y*-axis with the corresponding nucleotide position listed on the *x*-axis. Dark blue, dark red and green lines represent data from unmodified, ^{me}U and PU targets respectively. Rolling average percentages (9 nt window size) of array A·T and A·G content are plotted in light blue and pink respectively relative to the right *y*-axis.

changes not involving insertions and deletions greater than the probe length. Every *BRCA1* nucleotide position is interrogated by four 25 nt probes on the chip, each substituted with one of the four nucleotides in the central position. The ratio of hybridization signal to perfect match probes relative to those of the three single nucleotide substitution mismatch probes provides a measurement of hybridization specificity under a given set of conditions (3,4,6,7). A set of perfect match oligonucleotides (25, 23 and 22 nt in length) per target strand base form a contig of single nucleotide overlapping probes which tile across all *BRCA1* coding exons along with 10 bp of flanking intronic sequences. These probes are used in two color hybridization mutational analysis experiments (6,7).

Modified nucleoside triphosphate design

To directly test the effects of target sequence composition due to dA·rU and dT·rA probe·target interactions on the range of hybridization signals, we incorporated modified nucleotides into *BRCA1* targets. Elegant studies have shown that several RNA polymerases can tolerate template-directed incorporation of non-natural nucleoside triphosphates (17,18). Furthermore, modified uridine derivatives have been characterized which enhance hybridization with adenosine-rich targets including 5-methyluridine (^{me}U) (19) and 5-(1-propynyl)-deoxyuridine (^{P}dU). The former is a naturally occurring post-transcriptional modification in several tRNA species (20) while the latter has been employed in antisense oligonucleotide gene expression inhibition studies (21). The enhanced thermodynamic stability of these modified uridine-containing base pairs, shown in Figure 1, has been postulated to be due to more favorable stacking interactions and entropic factors (21,22). These entropic factors may stem from the enhanced displacement of highly ordered water molecules from the duplex due to these modified uridines (21,22). 2,6-Diaminopurine (DAP) is a modified adenine base which enhances binding affinity to thymine although having significant affinity to other bases in some sequence contexts (23,24). This modified adenine has been proposed to increase the stability of base pairs with thymidine (Fig. 1) due to altered stacking interactions, the formation of an additional hydrogen bond in the modified base pair and removing the spine of hydration in the minor groove (23,24; Fig. 1).

Effects of modified nucleoside triphosphates on *in vitro* transcription reactions

In vitro transcription reactions were performed in the presence of ATP, CTP (including biotin-15-CTP for post-hybridization array staining with a phycoerythrin–streptavidin conjugate), GTP and UTP. When examining 5-methyl-UTP (^{me}UTP) and 5-(1-propynyl)-UTP (^{P}UTP) incorporation, UTP was excluded. Likewise, ATP was excluded when transcribing with diaminopurine riboside 5′-triphosphate (DTP). For T3 and T7 RNA polymerase *in vitro* transcription reactions, ^{me}UTP did not significantly effect transcription yield relative to UTP based on ethidium bromide staining of the 3.43 kb exon 11 transcription products on agarose gels (data not shown). Substitution of ^{P}UTP for UTP in T3 RNA polymerase-mediated transcription reactions caused an ~4-fold decrease in ethidium bromide stained reaction exon 11 transcription product. In the analogous T7 RNA polymerase-mediated reaction, an ~10-fold decrease in exon 11 transcription product was found. DTP caused a more dramatic decrease in exon 11 transcription product, 10- and 20-fold for T3- and T7-mediated transcription reactions, respectively. All reactions containing both DTP and ^{me}UTP or ^{P}UTP showed greatly diminished transcription product yields (>50-fold).

Equivolume amounts of transcription products were fragmented and diluted into hybridization buffer (Materials and Methods). Target concentrations were not adjusted since it is not cost practical to increase transcription reaction volumes for ^{P}UTP and DTP *in vitro* transcription reactions. Furthermore, co-transcription of the smaller exons increases sample throughput and decreases reagent usage but produces a mixture of RNA species which are difficult to resolve and quantitate by gel electrophoresis. In theory

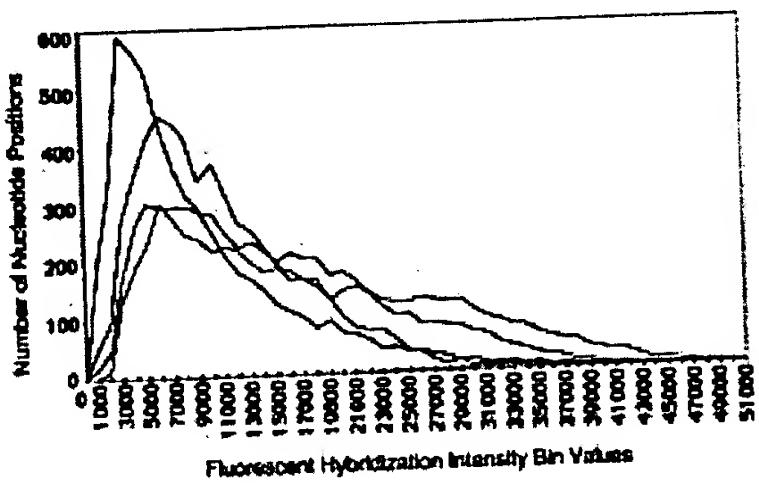


Figure 3. Binned hybridization signal intensities of modified antisense targets. Averaged fluorescent hybridization signal intensities of antisense targets (two experiments) to perfect match probes (buffer A, 40°C) are calculated and placed into bin values listed on the *x*-axis. The number of perfect match probes within these bin values (each encompassing 1000 fluorescence intensity units greater than listed value, except for the first two bins which correspond to perfect match probes having between 0 and 500 and those having between 500 and 1000 fluorescence intensity units) is shown on the *y*-axis. Dark blue, dark red, green and black lines represent data acquired from unmodified, ^{me}U, ^{P}U and DAP targets respectively.

it is possible to normalize the concentration of unmodified targets relative to the 5-(1-propynyl)-uridine and 2,6-diaminopurine modified targets. However, this would globally reduce hybridization for unmodified targets, making it difficult to accurately quantitate signals from all oligonucleotide probes. Therefore, the results generated from 5-(1-propynyl)-uridine and 2,6-diaminopurine (and to a lesser extent 5-methyluridine) containing targets should be taken as qualitative rather than quantitative. Our goal was to elucidate how modified nucleoside triphosphate incorporation would affect the performance of *BRCA1* target hybridization within the context of previously established assay conditions (7).

Hybridization properties of unmodified targets

The intensity of specific target hybridization may be shown by plotting hybridization signal strength to each perfect match probe per nucleotide position (Fig. 2). The fluorescent hybridization signal of perfect match probes varies over 130- and 230-fold (sense strand data) and 250- and 620-fold (antisense strand data) for unmodified target at 40°C in buffers A and B, respectively (based on averages of the 10 highest and lowest hybridization signals). Localized decreases in hybridization signal cannot be fully accounted for by thermodynamic parameters based upon target A/U or pyrimidine content (Fig. 2) but presumably also reflect potential intra- and intermolecular target and/or probe structures that inhibit hybridization (10). Of note is the use of tetramethylammonium (TMA) salts (25) which, along with betaine (26), have been widely used to minimize differences in oligonucleotide hybridization due to A·T content. These effects have been attributed to a non-cooperative differential stabilization of A·T (or A·U) base pairs, relative to G·C base pairs within duplex nucleic acid (25). Nevertheless, these buffers do not completely ameliorate energetic differences in the hybridization of short oligonucleotide targets (27). Although TMA^+ counterions altered hybridization to subsets of arrayed oligonucleotides relative to Na^+ counterions, they did not produce a globally uniform hybridization pattern (data not shown).

Table 2. Sequence tracts with enhanced hybridization due to analog incorporation

Sub[1]	Strand[2]	Tract[3]	Sequence[4]	Gain[5]
^{me}U	S	3062-3091	ACCACTTTTTCCCATCAAGTTCATTTTGTT	7.2
	S	3488-3509	cagaTTTCTCTCCATATCTGATTTCAgata	6.5
	S	3799-3808	ccctgcttccAACACTTGTTatttggtaaa	5.6
	S	2984-2988	ggttttgtctatcATCTCagttcagaggca	4.8
	S	1652-1665	acatcaggCCTTCATCCTGAGGattttatc	4.5
^{P}U	S	380- 408	tGAAATCATTTGTGCTTTTCAGCTTGACAc	8.8
	S	3491-3508	agatttCTCTCCATATCTGATTTCagataa	4.3
	S	355- 364	agtacgagatTTAGTCAACTtgttgaagag	3.3
	S	4828-4857	TCTTCTCTGATGACCCTGAATCTGATCCTT	3.2
	S	5531-5543	cacaggtgtCCACCCAATTGTGgttgtgca	2.9
DAP	S	826- 849	ataCTGAACATCATCAACCCAGTAATAatg	7.3
	S	3213-3220	taataacattAGAGAAAAtgtttttaaaga	6.9
	S	997-1003	gtttattactcaCTAAAGAcagaatgaatg	6.2
	S	1618-1623	gtcccctcacaaATAAATtaaagcgtaaaa	5.9
	S	1047-1053	ataCTGAACATCATCAACCCAGTAATAatg	5.6
^{me}U	AS	2309-2322	agagaagaAAAAGAAGAGAAACtagaaaca	20.1
	AS	4154-4164	cagatgatgaAGAAAGAGGAAcgggcttga	11.3
	AS	3448-3460	aaataaaaaAGCAAGAATATGAAgaagtag	10.9
	AS	1626-1632	acaaataaattaAAGCGTAaaaggagacct	10.9
	AS	2071-2076	gtgaagagataaAGAAAAaaaagtacaaca	9.7
^{P}U	AS	2309-2321	agagaagaAAAAGAAGAGAAActagaaaca	174.0
	AS	3160-3165	ctgaaagagaaaTGGGAAatgagaacattc	35.0
	AS	3216-3224	ataacattagaGAAAATGTTTttaaagaag	30.5
	AS	4155-4161	tcagatgatgaaGAAAGAGgaacgggcttg	28.6
	AS	3450-3458	aaataaaaaagCAAGAATATgaagaagtag	27.7
DAP	AS	3634-3655	taaggAAAGTTCTGCTGTTTTTAGCAAaag	5.0
	AS	3893-3912	gaggaGAATTTATTATCATTGAAGAatagc	4.6
	AS	254- 263	*TAATTTATAGTTTTGCATGct*gaaactt	4.3
	AS	3204-3233	AATAACATTAGAGAAAATGTTTTTAAAGAA	4.1
	AS	1433-1450	tcatgaGGCTTTAATATGTAAAAGtgaaag	4.0

[1]Target substitutions with ^{me}U, 5-methyluridine and ^{P}U, 5-(1-propynyl)-uridine.
[2]S, sense and AS, antisense strand data.
[3]Exonic nucleotide tracts with peak hybridization enhancements relative to unmodified targets (40°C, Buffer A).
[4]Tracts (*BRCA1* sense strand nucleotide sequence shown) displaying the five largest hybridization signal increases relative to unmodified targets. Italicized letters indicate intronic sequence and uppercase letters represent nucleotides with highest enhancement levels.
[5]Maximum level of hybridization signal enhancement.

Hybridization properties of modified targets

The range of fluorescent signal intensities was narrowed somewhat by ^{me}U and ^{P}U incorporation (Fig. 2). Global effects on the hybridization signal strength of ^{me}U, ^{P}U and DAP modified targets were further characterized by calculating the number of nucleotide positions having signal intensities within categorized bins (Fig. 3). Relative to unmodified antisense target, ^{me}U- and ^{P}U-substituted targets showed hybridization signals shifted towards overall higher values while those containing DAP shifted towards lower values. Relative to unmodified sense strand target, ^{me}U- and DAP-substituted targets showed similar average hybridization signal intensities (differing <1.2-fold), however, ^{P}U-substituted target had an ~2-fold decreased hybridization signal (buffer A, 40°C). This may result from lowered transcription reaction yields as well as from intra- and/or intermolecular target structures.

To assess localized changes in hybridization signal due to modified nucleotide incorporation, we calculated the ratio of perfect match probe hybridization signal intensity of modified to

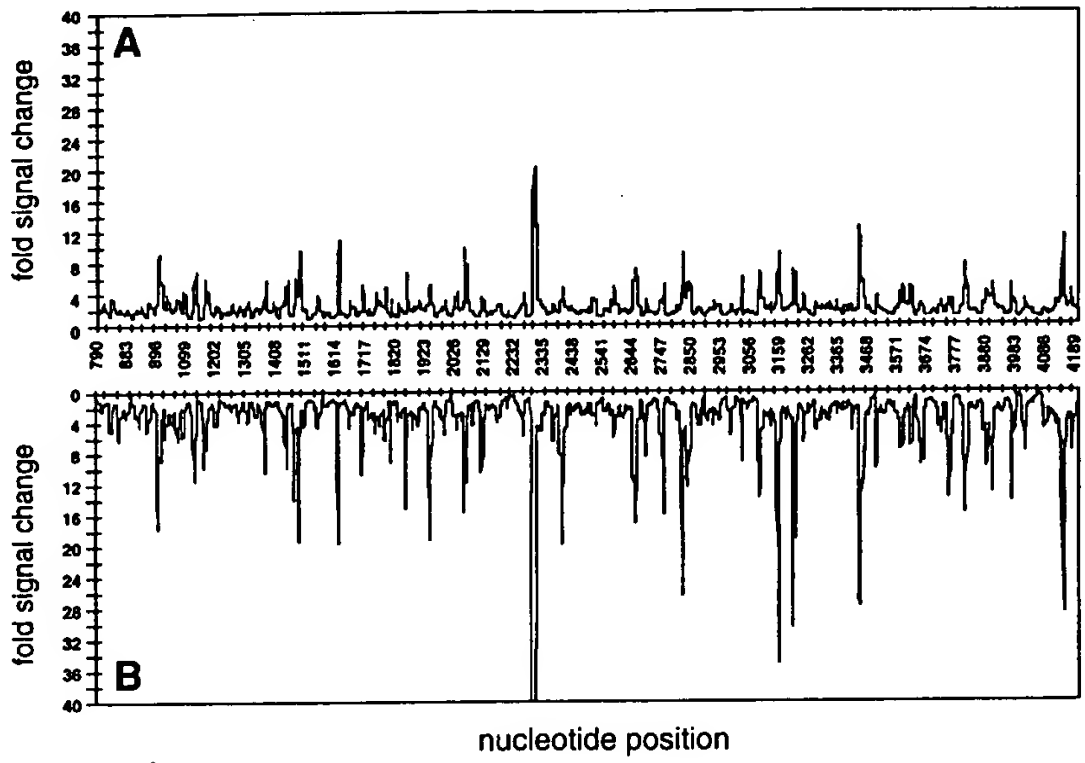


Figure 4. Relative hybridization intensities of modified pyrimidine antisense targets. Average fluorescence intensities of targets (two experiments) hybridized (buffer A, 40°C) to perfect match probes corresponding to exon 11 of the *BRCA1* oligonucleotide arrays were quantitated. Ratio of perfect match probe hybridization signals of (**A**) ^{me}U target relative to unmodified target and (**B**) PU target relative to unmodified target are given.

unmodified target for each sense and antisense strand nucleotide (Fig. 4). The largest signal enhancements in ^{me}U- and PU-substituted targets were primarily found in pyrimidine-enriched tracts containing a large number of uridine residues (Table 2). Homopyrimidine tracts are preferentially localized on the antisense *BRCA1* strand with 15 sequence tracts ≥10 nt long and only one such homopurine tract on this strand. Since DNA·RNA hybrid duplexes containing homopyrimidine RNA tracts are less stable than hybrid duplexes of identical sequence containing homopurine RNA tracts (28), modified uridine analogs have the best opportunity to significantly affect hybridization in these sequence contexts (Fig. 2). This could explain why ^{me}U and PU substitutions have a greater positive impact on hybridization signal from antisense than from sense target strands. While DAP-substituted targets also show regions of enhanced signal (Table 2) they are significantly less pronounced than the highest found in the pyrimidine-substituted targets. This may reflect the relative stability of DNA·RNA hybrids containing unmodified homopurine RNA strands.

Unmodified and ^{me}U-substituted targets show similar single nucleotide mismatch destabilization properties on both strands under a variety of assay conditions (Table 3 and Figs 5 and 6). Significantly, this selectivity is maintained with the antisense strand where ^{me}U incorporation produced the greatest localized hybridization signal enhancements (Fig. 4 and Table 2). When comparing hybridization properties of ^{me}U-substituted (buffer A, 42°C) relative to unmodified (buffer A, 40°C) targets, there was only a 2.3-fold average decrease in single nucleotide mismatch discrimination in the five antisense strand nucleotide tracts (Table 2) with the highest levels of signal enhancement. Lower temperatures are used in this comparison for unmodified targets due to diminished, thus less reliable, hybridization signal in these sequence tracts at 42°C.

The global single nucleotide mismatch specificity of probe·target interactions decreases when comparing PU-substituted (42°C) and unmodified targets under identical hybridization conditions (Table 3 and Fig. 5). This decreased specificity is highlighted in areas of the greatest probability of PU incorporation. When comparing hybridization properties of PU-substituted (buffer A, 42°C) relative to unmodified (buffer A, 40°C) targets, there was a 5.1-fold average decrease in single nucleotide mismatch discrimination in the five antisense strand nucleotide tracts (Table 2) with the highest levels of signal enhancement.

An example of the localized effects of modified pyrimidine substitutions on target hybridization signal strength from perfect match and single nucleotide mismatch probes is shown in Figure 6. While ^{me}U incorporation enhances hybridization signal and maintains a specific hybridization pattern to match probes in both sequence contexts, PU incorporation increases hybridization signal at the expense of single nucleotide mismatch specificity (Fig. 6c and f). Importantly, significantly increased PU- substituted target cross-hybridization often occurs in all four probes interrogating a single target nucleotide. Increased cross-hybridization to other areas of the array is also found with PU-substituted relative to unmodified oligodeoxyribonucleotides when hybridized in buffer A (data not shown).

DAP-substituted sense targets produce significantly stronger cross-hybridization to single base pair mismatch probes than unmodified targets (Table 3). Nevertheless, when comparing hybridization properties of DAP-substituted (buffer A, 42°C) relative to unmodified (buffer A, 40°C) targets, there was a only

Table 3. Single nucleotide mismatch specificity ratios

		unmodified				5-methyluridine				5-(1-propynyl)-uridine				2,6-diaminopurine		
	Temp °C[1]	35	40	42	45	35	40	42	45	35	40	42	45	35	40	42
	Bin[2]	% Frequency[3]				% Frequency				% Frequency				% Frequency		
Buffer A Sense Data	< 1.2	1.4	0.3	0.4	0.3	1.5	0.5	0.4	0.4	7.2	3.5	2.6	1.5	20.4	13.9	17.8
	1.2 - 2.0	25.5	16.0	13.8	6.5	32.0	22.9	21.3	18.3	35.2	29.1	21.4	16.5	43.3	45.0	34.9
	2.0 - 3.0	30.7	30.3	32.0	24.8	36.4	39.2	37.6	41.5	33.2	39.0	41.7	43.9	20.9	22.8	25.4
	3.0 - 4.0	17.6	20.5	23.3	25.5	16.3	19.8	21.0	22.7	14.3	16.8	22.0	24.1	8.0	8.8	11.1
	4.0 - 5.0	10.5	13.1	13.5	17.5	7.0	8.8	10.4	10.1	5.5	6.6	7.6	8.7	3.8	4.6	5.3
	5.0 - 6.0	5.7	7.6	7.3	10.6	3.2	4.3	4.5	3.8	2.4	2.6	2.7	2.8	1.6	2.3	2.8
	6.0 - 7.0	3.4	4.2	4.0	6.1	1.5	2.1	2.2	1.6	1.1	1.2	1.3	1.3	0.8	1.2	1.3
	7.0 - 8.0	2.1	3.1	2.1	3.3	0.9	1.0	1.2	0.8	0.5	0.5	0.5	0.7	0.4	0.6	0.6
	8.0 - 9.0	1.0	1.7	1.3	2.3	0.5	0.6	0.6	0.3	0.3	0.3	0.2	0.2	0.2	0.4	0.4
	9.0 - 10.0	0.7	0.9	0.6	1.1	0.2	0.3	0.4	0.2	0.1	0.2	0.1	0.2	0.1	0.1	0.2
	> 10.0	1.5	2.3	1.7	2.2	0.5	0.7	0.5	0.3	0.2	0.3	0.1	0.1	0.4	0.3	0.3
	Bin	% Frequency				% Frequency				% Frequency				% Frequency		
Buffer B Sense Data	< 1.2	3.6	1.1	1.0	0.9	3.2	1.3	0.6	0.5	6.4	1.6	1.2	1.1	46.4	22.3	19.2
	1.2 - 2.0	38.8	26.1	24.7	24.7	38.6	34.5	19.3	31.0	39.3	26.5	26.1	20.9	39.2	49.1	50.1
	2.0 - 3.0	28.0	31.9	32.1	32.2	30.6	34.3	34.6	37.9	32.3	38.4	37.6	41.4	10.3	18.9	19.6
	3.0 - 4.0	13.5	18.8	19.1	19.3	14.3	15.8	21.6	16.9	13.0	18.8	18.6	20.9	2.7	6.2	6.5
	4.0 - 5.0	7.1	10.7	11.0	10.7	6.3	7.3	12.1	7.2	5.3	8.0	8.3	9.1	0.8	1.9	2.6
	5.0 - 6.0	3.4	5.4	5.8	5.6	3.3	3.2	5.6	3.2	2.1	3.6	4.2	3.9	0.5	0.9	1.0
	6.0 - 7.0	2.0	2.8	2.8	2.9	1.6	1.8	2.8	1.5	0.9	1.6	1.9	1.4	0.1	0.4	0.6
	7.0 - 8.0	1.3	1.5	1.5	1.5	1.0	0.8	1.5	0.8	0.4	0.7	0.9	0.8	0.1	0.2	0.2
	8.0 - 9.0	0.7	0.7	0.8	1.0	0.6	0.5	0.8	0.5	0.2	0.4	0.5	0.3	0.0	0.1	0.1
	9.0 - 10.0	0.4	0.4	0.5	0.6	0.3	0.3	0.5	0.2	0.1	0.2	0.3	0.1	0.0	0.0	0.1
	> 10.0	1.1	0.7	0.6	0.7	0.3	0.3	0.8	0.2	0.1	0.2	0.3	0.1	0.0	0.0	0.1
	Bin	% Frequency				% Frequency				% Frequency				% Frequency		
Buffer A Antisense Data	< 1.2	0.8	0.2	0.3	0.5	1.1	0.1	0.2	0.3	9.3	3.8	3.1	1.1	2.9	0.9	0.5
	1.2 - 2.0	16.6	7.4	4.3	4.0	17.6	11.5	7.5	7.1	40.5	32.0	27.5	16.8	24.0	14.8	15.1
	2.0 - 3.0	28.8	20.8	14.9	17.2	33.1	31.2	28.0	29.7	34.4	37.1	39.0	39.1	30.6	26.3	33.2
	3.0 - 4.0	19.2	21.1	20.9	26.6	23.2	24.2	24.1	28.5	10.8	16.0	18.4	24.7	18.7	20.7	22.2
	4.0 - 5.0	12.1	15.6	17.2	20.8	11.8	13.4	16.4	17.0	3.2	6.1	7.0	11.0	10.0	13.5	11.7
	5.0 - 6.0	7.4	10.4	12.8	12.5	6.0	7.5	8.8	8.7	1.0	2.6	2.7	4.3	5.1	8.3	6.7
	6.0 - 7.0	4.2	7.3	8.7	7.2	3.0	4.7	5.7	4.4	0.5	1.1	1.2	1.7	3.0	5.4	4.0
	7.0 - 8.0	3.3	5.2	6.0	4.2	1.4	2.7	3.3	2.2	0.2	0.6	0.5	0.8	1.8	3.5	2.2
	8.0 - 9.0	2.0	3.1	4.1	2.6	1.0	1.6	2.2	1.2	0.1	0.2	0.2	0.3	1.1	2.2	1.2
	9.0 - 10.0	1.4	2.4	2.9	1.7	0.6	1.0	1.2	0.4	0.0	0.2	0.1	0.1	0.7	1.3	0.9
	> 10.0	4.2	6.7	7.9	2.8	1.1	2.0	2.5	0.5	0.1	0.3	0.2	0.2	2.1	3.2	2.2
	Bin	% Frequency				% Frequency				% Frequency				% Frequency		
Buffer B Antisense Data	< 1.2	1.7	1.6	1.8	2.6	0.9	0.3	0.2	0.4	3.2	0.7	0.5	0.5	7.2	3.9	2.8
	1.2 - 2.0	14.1	13.3	11.2	15.2	16.3	13.0	12.7	12.8	35.1	17.4	17.5	21.4	32.5	26.7	24.2
	2.0 - 3.0	26.2	25.6	23.6	29.8	31.4	31.8	33.0	31.5	37.7	36.0	36.7	41.4	28.0	29.8	30.4
	3.0 - 4.0	19.4	19.6	20.2	22.5	21.7	22.5	22.6	23.6	14.1	21.9	20.9	19.4	14.0	16.1	17.2
	4.0 - 5.0	12.5	12.9	13.9	13.2	11.6	13.3	13.5	14.4	5.7	10.9	11.1	9.1	7.6	8.4	9.6
	5.0 - 6.0	8.1	8.6	9.2	7.4	6.9	7.6	7.2	8.4	2.1	5.3	5.5	4.5	3.6	4.9	5.2
	6.0 - 7.0	5.4	5.6	6.0	4.2	4.0	4.5	4.5	3.8	0.9	3.0	3.3	1.9	2.3	3.0	3.4
	7.0 - 8.0	3.5	3.6	4.1	2.3	2.5	2.6	2.4	2.1	0.5	1.9	1.9	0.9	1.5	2.2	2.2
	8.0 - 9.0	2.6	2.5	3.0	1.3	1.7	1.7	1.6	1.2	0.2	1.1	1.1	0.5	1.0	1.3	1.6
	9.0 - 10.0	1.8	1.9	1.9	0.6	1.0	1.0	0.7	0.8	0.1	0.5	0.6	0.2	0.7	0.9	1.0
	> 10.0	4.8	4.8	5.0	0.8	1.9	1.7	1.5	1.1	0.3	1.1	0.9	0.2	1.7	2.8	2.3

[1]Hybridization temperature.
[2]Single nucleotide mismatch specificity ratio bins.
[3]% of BRCA1 coding nucleotide positions within the specified ratio bin.

a 2.3-fold average decrease in single nucleotide mismatch discrimination in the five sense strand nucleotide tracts (Table 2) with the highest levels of signal enhancement. Therefore, cross-hybridization occurs in different sequence contexts distributed throughout the array. Furthermore, DAP-substituted antisense targets showed single nucleotide mismatch specificities similar to unmodified targets, presumably due to the decreased number of adenine-rich tracts on this strand and thus lower levels of DAP incorporation. For both sense and antisense analysis, lower overall transcriptional yields of DAP-modified targets result in a lower target concentration in the hybridization reaction. This increases the stringency of the hybridization reaction and consequently increases single nucleotide mismatch discrimination. Increasing the DAP-modified target concentration to produce more robust hybridization signals, especially for antisense strand targets, will result in lower single nucleotide discrimination, thus reducing the usefulness of this modification.

Co-incorporation of DAP and modified uridine residues into sense and antisense targets had a significantly negative effect on hybridization specificity (data not shown). A large component of this problem is the decreased transcription product yield (>50-fold). It does not appear feasible to simultaneously incorporate modified pyrimidine and purine nucleotide triphosphates and retain robust product yield which has non-degenerate hybridization specificity with these enzymes under these conditions. In the future, RNA polymerases may be engineered to have increased ability to

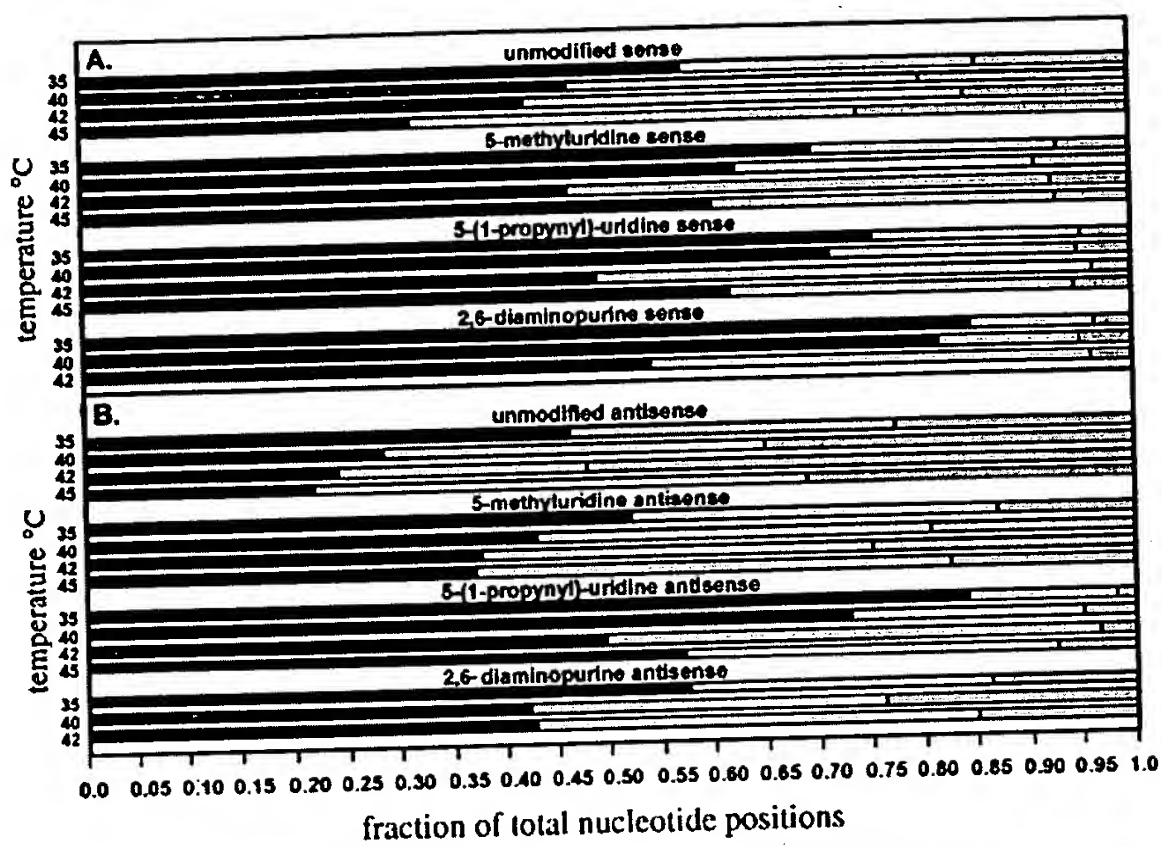


Figure 5. Single nucleotide mismatch specificity ratios of modified targets. Averaged ratios of target hybridization signals from perfect match probe to next highest mismatch probe taken from a minimum of two experiments (buffer A, 40°C). The *x*-axis represents the fraction of nucleotide positions within categorized bins of single nucleotide mismatch ratios <1.2 (red bars), between 1.2 and 3 (blue bars), between 3 and 5 (yellow bars) and >5 (orange bars), respectively. Sense and antisense strand data are shown in (**A**) and (**B**), respectively. Selected data of the single nucleotide mismatch ratios <1.2 are shown on a magnified *x*-axis in (**C**).

incorporate modified nucleoside triphosphates into transcription products.

Potential applications for modified targets

The examined modified triphosphates can be incorporated into large RNA transcripts by T3 and T7 RNA polymerases. While DAP and PU incorporation both lead to enhanced hybridization in specific sequence contexts, the loss of binding specificity reduces the likelihood of their use in mutation screening analysis. ^{me}U incorporation enhances target hybridization signals within specific sequence contexts and does not substantially increase hybridization signal to single nucleotide mismatch probes. Enhanced ^{me}U RNA target hybridization signals will be especially important in the mutational analysis of genes having localized regions of strongly biased A·T sequence content (i.e. *BRCA2* and *ATM* genes). Furthermore, in conjunction with modified oligonucleotide surface probes (Fidanza *et al.*, unpublished observations) it may be possible to normalize the binding affinity of all perfect match probes in an array (29). This would potentially allow hybridization conditions to be used which globally optimize hybridization signal strength and specificity.

Modified nucleoside triphosphate usage could also benefit RNA expression monitoring experiments based on hybridization

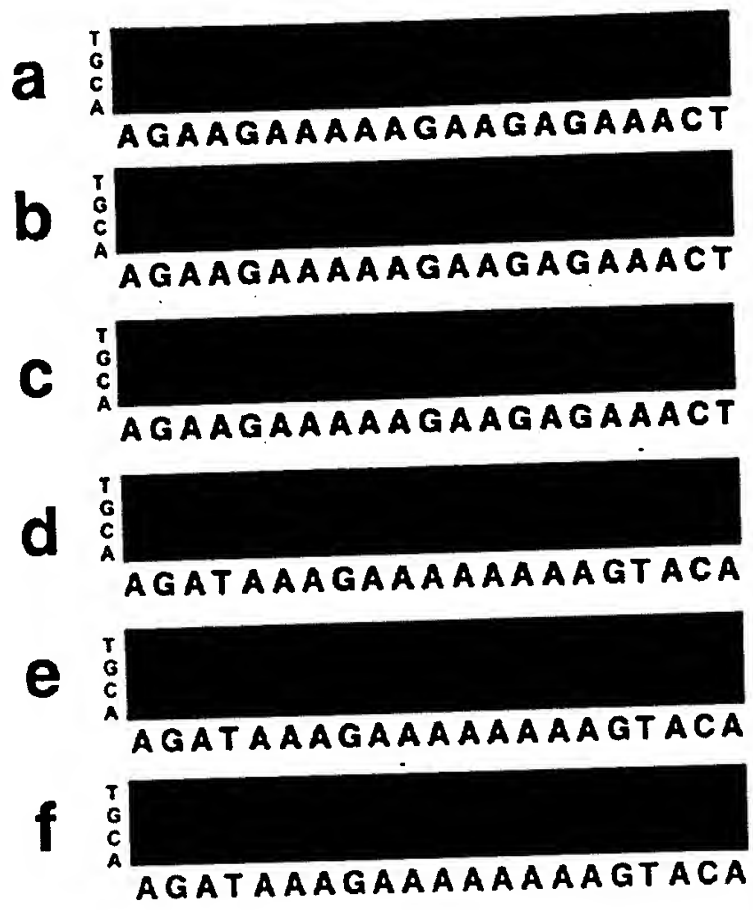


Figure 6. Modified pyrimidine target image comparisons. Magnified digitized false color images showing hybridization pattern of *BRCA1* antisense targets (buffer A, 40°C). Brightness and contrast settings are changed in each panel to increase image clarity. Nucleotide identities, determined through dideoxy-sequencing analysis, are given under the respective column and colored red or blue if correctly or incorrectly identified by hybridization analysis (perfect match probe intensity being at least 1.2-fold greater than that of the next highest mismatch probe intensity), respectively. Several base calls may be difficult to visualize due to limitations in printing technology as well as in the linear range of the human eye for detecting monochromatic color changes. (**a**–**c**) Hybridization patterns of unmodified, ^{me}U and PU antisense strand targets, respectively, to nt 2303–2323 of *BRCA1* cDNA. (**d**–**f**) Hybridization patterns of unmodified, ^{me}U and PU antisense strand targets, respectively, to nt 2065–2085 of *BRCA1* cDNA.

to high density oligonucleotide arrays (8). In such experiments, perfect match probe oligonucleotides are selected based upon sequence composition effects to produce robust and specific hybridization signals from RNA targets (8). Targets containing modified bases may have increased affinity towards a number of perfect match probes previously giving a poor hybridization signal. This would expand the variety of oligonucleotide probes which could be used in these experiments and allow increased freedom in selecting probes placed in strategic positions (i.e. splice junction sequences so as to monitor the expression of differentially spliced RNA transcripts).

Modified RNA transcripts can also be used to analyze the biophysical properties of nucleic acid structures (17,18,30). Others have incorporated modified bases into RNA during *in vitro* selection assays to expand the repertoire of RNA species having affinities for small molecules or specific catalytic properties (31). Increasing the stability of A·U base pairs using this strategy could expand the variety of nucleic acid structures which may have distinct biophysical properties.

ACKNOWLEDGEMENTS

We would like to thank Larry Brody and Bryan Sun for thoughtful discussion. These studies have been supported in part by research grant 5POLHGO1323-03 to SPAF.

REFERENCES

1 Fodor,S.P.A., Read,J.L., Pirrung,M.C., Stryer,L., Lu,A.T. and Solas,D. (1991) *Science*, **251**, 767–773.
2 McGall,G., Labadie,J., Brock,P., Wallraff,G., Nguyen,T. and Hinsberg,W. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 13555–13560.
3 Cronin,M.T., Fucini,R.V., Kim,S.M., Masino,R.S., Wespi,R.M. and Miyada,C.G. (1996) *Hum. Mutat.*, **7**, 244–255.
4 Kozal,M.J., Shah,N., Yang,R., Fucini,R., Merigan,T.C., Richman,D.D., Morris,D., Hubbell,E. and Gingeras,T.R. (1996) *Nature Med.*, **2**, 753–759.
5 Yershov,G., Barsky,R., Belgovskiy,A., Kirillov,E., Kreindlin, Ivanov,I., Parinov,S., Guschin,D., Drobishev,A., Dubiley,S. and Mirzabekov,A. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 4913–4918.
6 Chee,M., Yang,R., Hubbell,E., Huang,X.C., Stern,D., Winkler,J., Lockhart,D.J., Morris,M.S. and Fodor,S.P. (1996) *Science*, **274**, 610–614.
7 Hacia,J.G., Brody,L.C., Chee,M.S., Fodor,S.P.A. and Collins,F.S. (1996) *Nature Genet.*, **14**, 441–447.
8 Lockhart,D.J., Dong,H., Byrne,M.C., Follettie,M.T., Gallo,M.V., Chee,M.S., Mittmann,K.M., Wang,C., Kobayashi,M., Horton,H. and Brown,E.L. (1996) *Nature Biotech.*, **14**, 1675–1680.
9 Shoemaker,D.D., Lashkari,D.A., Morris,D., Mittmann,M. and Davis,R.W. (1996) *Nature Genet.*, **14**, 450–456.
10 Milner,N., Mir,K.U. and Southern,E.M. (1997) *Nature Biotech.*, **15**, 537–541.
11 Hacia,J.G., Makalowski,W., Edgemon,K., Erdos,M.R., Robbins,C.M., Fodor,S.P.A., Brody,L.C. and Collins,F.S. (1998) *Nature Genet.*, **18**, 155-158.
12 Hobbs,F.W. (1989) *J. Org. Chem.*, **54**, 3420–3422.
13 Sowa,T. and Ouchi,S. (1975) *Bull. Chem. Soc. Jpn*, **48**, 2084–2090.
14 Ludwig,J. (1981) *Acta Biochim. Biophys. Acad. Sci. Hung.*, **16**, 131–133.
15 Howard,F.B. and Miles,H.T. (1984) *Biochemistry*, **23**, 6723–6732.
16 Moffatt,J.G. and Khorana,H.G. (1961) *J. Am. Chem. Soc.*, **83**, 649–658.
17 Piccirilli,J.A., Krautch,T., Moroney,S.E. and Benner,S.A. (1990) *Nature*, **343**, 33–37.
18 Tor,Y. and Dervan,P.B. (1993) *J. Am. Chem. Soc.*, **115**, 4461–4467.
19 Zimmer,M. and Scheit,K.H. (1984) *Nucleic Acids Res.*, **12**, 2243–2258.
20 Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York, NY.
21 Froehler,B.C., Wadwani,S., Terhorst,T.J. and Gerrard,S.R. (1992) *Tetrahedron Lett.*, **33**, 5307–5310.
22 Guitierrez,A.J., Matteucci,M.D., Grant,D., Matsumura,S., Wagner,R.W. and Froehler,B.C. (1997) *Biochemistry*, **36**, 743–748.
23 Chollet,A. and Kawashima,E. (1988) *Nucleic Acids Res.*, **16**, 305–317.
24 Chaejoon,C., Tinoco,I. and Chollet,A. (1988) *Nucleic Acids Res.*, **16**, 5115- 5122.
25 Melchior,W.B. and von Hippel,P.H. (1973) *Proc. Natl Acad. Sci. USA*, **70**, 298–302.
26 Rees,W.A., Yager,T.D., Korte,J. and von Hippel,P.H. (1993) *Biochemistry*, **32**, 137–144.
27 Ricelli,P.V. and Benight,A.S. (1993) *Nucleic Acids Res.*, **21**, 3785–3788.
28 Ratmeyer,L., Vianyak,R., Zhong,Y.Y., Zon,G. and Wilson,W.D. (1994) *Biochemistry*, **33**, 5298–5304.
29 Hoheisel,J.D. (1996) *Nucleic Acids Res.*, **24**, 430–432.
30 Moriyama,K., Negishi,K., Briggs,M.S.J., Smith,C.L., Hill,F., Churcher,M.J., Brown,D.M. and Loakes,D. (1998) *Nucleic Acids Res.*, **26**, 2105–2111.
31 Tarasow,T.M., Tarasow,S.L. and Eaton,B.E. (1997) *Nature*, **389**, 54–57.

BMC Biotechnology

BioMed Central

Research article

Quantitative assessment of the use of modified nucleoside triphosphates in expression profiling: differential effects on signal intensities and impacts on expression ratios

Allen Nguyen, Connie Zhao, David Dorris and Abhijit Mazumder*

Address: Motorola Life Sciences, 4088 Commercial Avenue, Northbrook, Illinois 60062, USA

E-mail: Allen Nguyen - allenn1999@yahoo.com; Connie Zhao - acz020@email.mot.com; David Dorris - David.Dorris@motorola.com; Abhijit Mazumder* - arm034@email.mot.com

*Corresponding author

Published: 31 July 2002

Received: 9 May 2002
Accepted: 31 July 2002

BMC Biotechnology 2002, 2:14

This article is available from: http://www.biomedcentral.com/1472-6750/2/14

Abstract

Background: The power of DNA microarrays derives from their ability to monitor the expression levels of many genes in parallel. One of the limitations of such powerful analytical tools is the inability to detect certain transcripts in the target sample because of artifacts caused by background noise or poor hybridization kinetics. The use of base-modified analogs of nucleoside triphosphates has been shown to increase complementary duplex stability in other applications, and here we attempted to enhance microarray hybridization signal across a wide range of sequences and expression levels by incorporating these nucleotides into labeled cRNA targets.

Results: RNA samples containing 2-aminoadenosine showed increases in signal intensity for a majority of the sequences. These results were similar, and additive, to those seen with an increase in the hybridization time. In contrast, 5-methyluridine and 5-methylcytidine decreased signal intensities. Hybridization specificity, as assessed by mismatch controls, was dependent on both target sequence and extent of substitution with the modified nucleotide. Concurrent incorporation of modified and unmodified ATP in a 1:1 ratio resulted in significantly greater numbers of above-threshold ratio calls across tissues, while preserving ratio integrity and reproducibility.

Conclusions: Incorporation of 2-aminoadenosine triphosphate into cRNA targets is a promising method for increasing signal detection in microarrays. Furthermore, this approach can be optimized to minimize impact on yield of amplified material and to increase the number of expression changes that can be detected.

Background

DNA microarrays have been widely adopted in genomics because of the their ability to simultaneously examine the expression levels of thousands of genes. As a result, the scope of applications for microarrays has broadened rapidly, from drug discovery [1], to classification of cancers [2–4] and analysis of splice variants [5]. Novel analytical tools have been constructed to address every component of the microarray experiment and optimize performance [6]. However, ideal systems with maximized sensitivity and data reproducibility have not been achieved. One approach to enhance sensitivity in microarrays, using a novel signal amplification technique, has recently been reported [7]. Another approach is to increase the affinity

of a probe (nucleic acid present on the array) for its target through modifications to the length [8], chemistry [9], or physical structure [10] of the probe.

Naturally occurring analogs of purine and pyrimidine bases have been examined extensively for their ability to increase the thermodynamic stability of DNA:DNA and DNA:RNA duplexes [11–15]. Among these are 2-aminoadenine also known as diaminopurine (DAP), which is found in S-2L cyanophage DNA [16], 5-methyl uracil (MeU), and 5-methyl cytidine (MeC). Previous studies have shown the ribonucleoside triphosphate derivatives of modified bases to be effectively incorporated by RNA polymerases [17,18], which make them an excellent substrate for use in microarray sample preparations requiring amplification through *in vitro* transcriptions (IVTs).

Due to their effects on duplex stability and secondary structures and their lack of replication by DNA polymerases, modified nucleotides have recently been exploited in a variety of technologies in molecular biology and genomics. For example, 2'-O-methyl ribonucleotides and 5-(1-propynyl)pyrimidines have been chemically incorporated into oligonucleotides which were used to detect telomeric repeat sequences in fluorescence *in situ* hybridization (FISH) assays [19]. Furthermore, chimeric primers containing deoxynucleotides and 2'-O-methyl ribonucleotides have been used to eliminate artifacts, produced by exponential amplification of minor side-products, in cycle sequencing [20].

Recently, studies employing DNA arrays have also examined the use of modified nucleotides such as DAP, 5-bromodeoxyuridine, and 2'-O-methylthymidine in the probe [21,22]. An alternative method to increase the probe-target affinity is to incorporate the modified ribonucleotide, as the triphosphate derivative, during the IVT process so that the cRNA produced has the desired level of substitution of the corresponding unmodified nucleotide. Such an approach has been demonstrated in a study which applied these modified cRNA products to high density oligonucleotide arrays [23]. However, that study measured the signal when amplifying a specific gene rather than a genome-wide approach. A genome-wide amplification, coupled with the incorporation of modified nucleotides, permits measurement of every transcript in the sample. To our knowledge, no systematic studies have been performed examining the incorporation of modified ribonucleotide triphosphates into cRNA, yield of amplified material, effects on hybridization intensities and reproducibility, and the impact on the differential expression ratios. It will not be possible to gauge the full potential, advantages, and disadvantages until such studies have been completed.

In this study, we investigated the possibility that incorporation of modified nucleotides into complementary RNA (cRNA) target samples could increase signal intensity on the Motorola Codelink™ Expression microarray platform. The ratios of modified to unmodified nucleoside triphosphates (NTPs) were varied in each target synthesis in order to measure the range of effects on cRNA yield, specific activity of target sample, hybridization signals, and differential expression ratios. Our results suggest that incorporation of 2-aminoadenosine (DAP) triphosphate into target cRNA samples may increase hybridization signal intensity for a wide variety of RNA:DNA hybrid duplexes. In contrast, 5-methyluridine and 5-methylcytidine had detrimental effects on signal intensities. Hybridization specificity was dependent on both target sequence and extent of substitution with the modified nucleotide. Concurrent incorporation of modified and unmodified ATP in a 1:1 ratio resulted in significantly greater numbers of above-threshold ratio calls across tissues, while preserving ratio integrity and reproducibility.

Results

Effect of modified nucleotides on cRNA yield and assessment of their incorporation and biotin-11-UTP incorporation

Although DAP, MeC, and MeU were tolerated by T7 RNA polymerase, incorporation of these analogs into our IVT reaction cocktails reduced the yields of amplified cRNA. The magnitude of the decrease was 20–30% and depended on RNA tissue source and the NTP modification. However, within the context of our pre-established assay conditions (a single color, single sample per array system using ten micrograms of cRNA), enough cRNA was generated from each IVT reaction to perform hybridizations in triplicate. Thus, we were able to normalize input amounts of cRNA for each condition tested in order to quantitatively compare relative changes due to each respective modification. In these experiments, we generated cRNA from five micrograms of total RNA and did not explore the impact on yield when smaller or larger amounts of input total RNA are used.

We next determined how well the modified NTPs were incorporated by the T7 RNA polymerase and whether incorporation of biotinylated UTP was altered due to the presence of these modified NTPs. To address these questions, we used an analytical method developed in our laboratory and described in a previous study [24]. Briefly, the complex cRNA is digested with P1 nuclease and calf intestinal phosphatase and applied to a high performance liquid chromatography (HPLC) column to separate the nucleosides, followed by an absorbance measurement at 260 nm. As seen in Figure 1A, when only the unmodified NTPs are incorporated during the IVT, there is good separation of the individual nucleosides. Furthermore, as re-

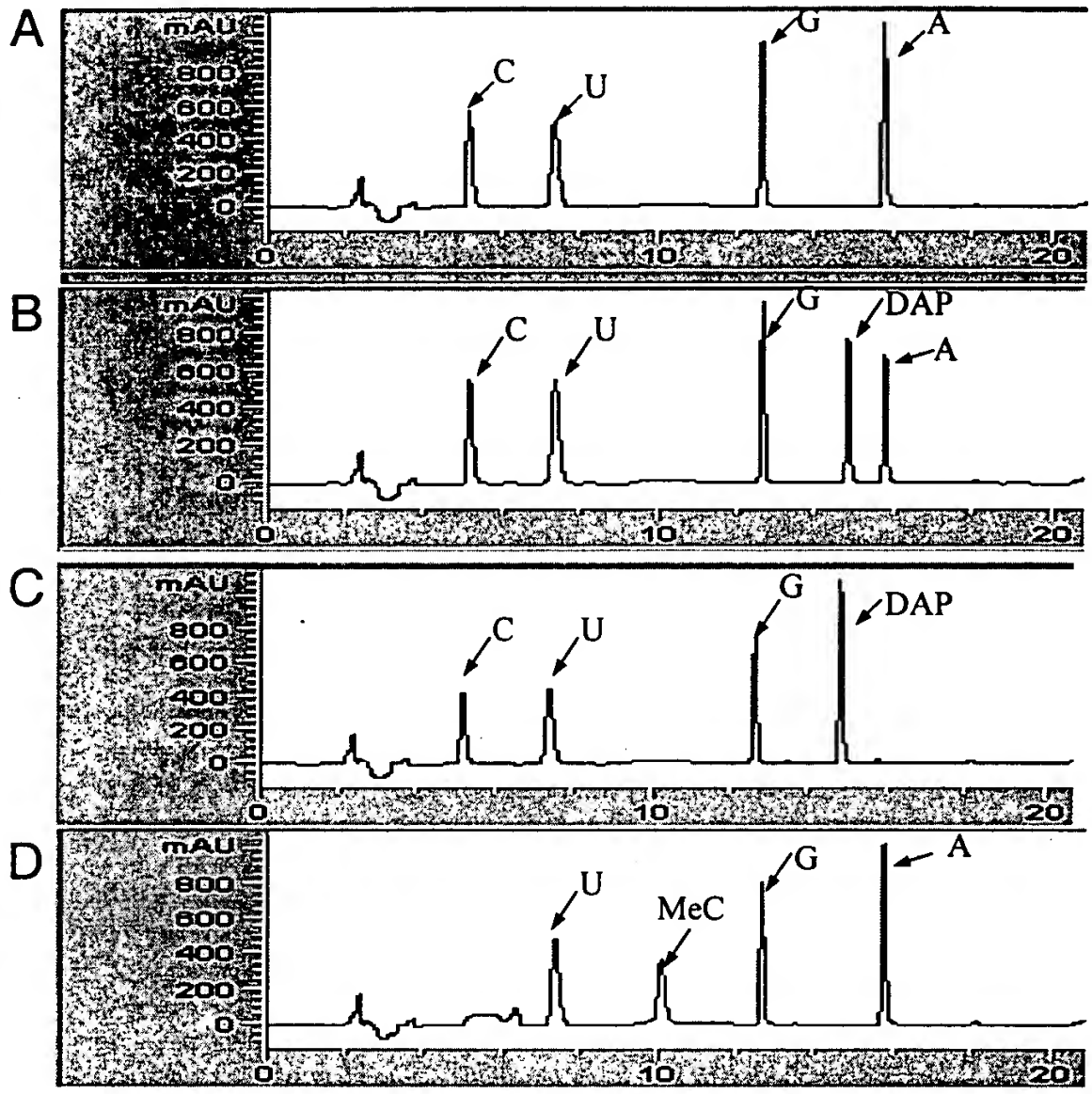


Figure 1
HPLC analysis of digested cRNA demonstrates incorporation of modified nucleotides and their resolution from unmodified counterparts. Absorbance profiles at 260 nm are shown for (A) control, (B) 1:1 DAP:A, (C) fully substituted DAP, and (D) fully substituted MeC conditions. Proportions of each nucleoside were calculated using peak areas and extinction coefficients. The peaks for cytosine, uridine, guanosine, and adenosine show up in the unmodified control sample (A) at approximately 5.0, 7.2, 12.6, and 15.6 min, respectively.

ported earlier [24], using the extinction coefficients and integrating the area under these peaks, there are approximately equal amounts of each of the four nucleosides. When DAP was added at a 1:1 molar ratio to the adenosine, the chromatogram showed incorporation of DAP into the cRNA, and this level of incorporation appeared to be equivalent to the level of incorporation of adenosine (Figure 1B). Moreover, when the adenosine was fully substituted by the DAP in the IVT reaction cocktail, only a peak corresponding to the DAP was detected (Figure 1C), and this peak was approximately equal in area to that of the adenosine in the control situation. Similarly, when MeC was fully substituted for cytidine in the IVT reaction cocktail, only a peak corresponding to the MeC was detected (Figure 1D), and this peak was approximately equal in area to that of the cytidine in the control situation. In contrast, we were unable to examine incorporation levels of the MeU because the MeU peak was eluted at approximately the same time as the guanosine peak (data not shown). It is important to note that these analyses used the complex cRNA and represent a global, average view of the incorporation. Incorporation may differ somewhat depending on the sequence, structure, or expression level of the nascent RNA transcript. We therefore incorporated each of these modified nucleotides into a unique bacterial transcript, generated by run off transcription from a plasmid into which the gene was cloned. When this transcript was digested and applied to the HPLC, we found similar results as seen with the complex cRNA (data not shown). We conclude that these modified nucleotides are incorpo-

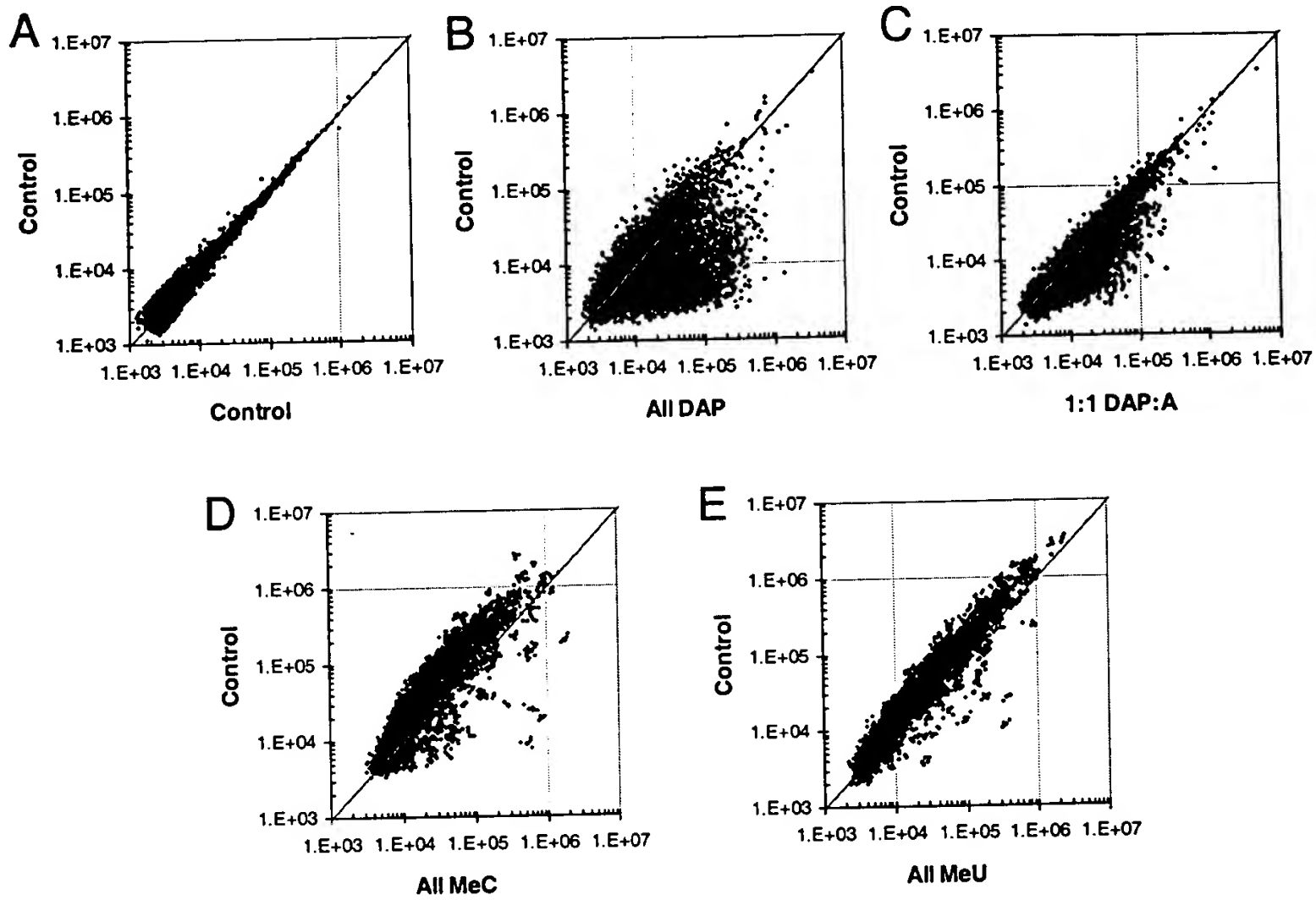


Figure 2
Modified purines and pyrimidines show differential effects on hybridization intensities when incorporated into cRNA targets. Modified and unmodified (Control) samples were normalized for concentration, fragmented, and hybridized onto Human Uniset I microarrays in duplicate, except for 5-methyl CTP samples, which were hybridized in duplicate onto arrays with 1100 probes spotted in 6-fold redundancy. Average signal intensities after background subtraction were plotted for all probes of the Human Uniset I array for the control samples versus (A) control, (B) fully DAP-substituted, (C) 1:1 substituted DAP, (D) fully meC-substituted, and (E) fully MeU-substituted.

rated during the IVT reaction at ratios which are reflective of the input ratio to the unmodified counterpart.

Because the sensitivity and reproducibility of microarrays depends on the specific activity of target cRNA, we also wanted to determine if adding modified NTPs to the IVT changed the incorporation rate of biotin-11-UTP. This biotinylated nucleotide is used in our biotin-streptavidin-Alexa647 conjugate detection system. We found that the level of biotinylated uridine, detected at 294 nm after digestion and application to the HPLC, was unaffected by the presence of any of the modified NTPs in the IVT reaction cocktail (data not shown). Moreover, analysis of individual transcripts, which were enzymatically digested into mononucleosides, showed equivalent incorporation rates of biotin-11-UTP for both control samples and target samples containing modified ATP.

Lastly, we examined what effect, if any, incorporation of modified NTPs had on the length of the amplified cRNA. Transcript size was determined by running samples on an Agilent 2100 BioAnalyzer. No difference was observed for either individual transcripts or complex samples in modified versus unmodified ATP samples (data not shown).

Effect of modified nucleotides on hybridization signals and specificity

We next determined the effect of these modified NTPs on hybridization intensities. After hybridization, scatter plots of hybridization intensities were generated comparing those intensities generated from unmodified cRNA (con-

Table 1: Percent signal change due to nucleotide substitution.

	Relative percent change		
Probe Name	MeC	MeU	1:1 DAP:A
AI936591	108	-38	27
AB029001	110	-7	112
U90544	113	-19	540
AA707570	115	-7	259
AK000445	116	22	23
AF054506	121	-36	47
AB007891	127	-16	117
D42045	138	62	35
AB006190	146	-28	481
AI935557	166	25	2
NM_006794	168	-20	1623
AB028946	168	-24	351
AK000601	179	51	451
AB020710	198	90	165
AK000527	230	-33	93
AI742085	246	5	41
AF017789	269	601	-44
AK001962	279	257	4
AF038661	291	-44	445
AB037797	301	295	279
AF035121	314	40	10
AF016369	332	324	149
AB034695	346	-22	363
U15932	377	66	43
AK001933	380	219	237
AJ223352	465	545	54
AF227899	559	106	603
AK000286	821	-57	922
AF052093	885	587	177
L77701	2356	1278	284
AL137493	4256	-37	2977
D86956	5999	2857	1367

MeC and MeU substitution increases signal intensities for a small number of probes. Relative percent change was calculated by the equation: 100 * (modified sample signal - control sample signal) / control sample signal. Relative change for these probes due to 1:1 DAP:A substitution is also given as reference.

trol) to either the duplicate control hybridization or to intensities generated from cRNA where one of the modified NTPs was incorporated. The comparison of control versus control (Figure 2A) illustrates that, if hybridization intensities are equal in the two conditions, the points should lie on the diagonal. Complete substitution of unmodified ATP with DAP (Figure 2B) produced greater increases overall in individual probe signals than the target sample containing a 1:1 ratio of DAP to unmodified ATP (Figure 2C), as would be expected if DAP incorporation increased the DNA:RNA hybrid duplex stability. A quantitative estimation of assay performance increase can be derived from the median probe signal on each microarray. Although Figure 2B and 2C show that increases in signal intensity are not linear for all probe sequences, median slide intensity across sets of duplicate hybridizations for two different tissues (human embryonic kidney and Burkitt's lymphoma) increased an average of 40% +/- 7% and 99% +/- 6% for the 1:1 DAP:A and all DAP conditions, respectively, over the unmodified control sample. However, Figure 2B also shows that full substitution of adenosine by DAP reduced the hybridization signals for many of the probes, generating a bowing towards the unmodified control condition and suggesting a duplex destabilizing effect of complete DAP substitution in certain sequence contexts [11,23]. On a global scale, such a high incidence of probes with reduced hybridization signals may limit the usefulness of the complete DAP substitution. Of particular interest is the observation that the number of probes which have increased signal and the degree to which their signals are increased are highest at the lower end of the signal range. In contrast, both of the modified pyrimidine triphosphates (MeC and MeU) tested resulted in a decrease in overall signal intensities for nearly all of the probes tested when compared against unmodified control samples (Figures 2D &2E). However, data points on the right of the diagonal in figures 2D and 2E also show that for a fractional set of target-probe pairs, substitution with MeC and MeU increased hybridization signal relative to the unmodified control. Those probes whose signal intensities were increased at least two-fold are summarized in Table 1. As Table 1 shows, for a number of transcripts substitution with MeC or MeU may potentially augment hybridization signals more so than even DAP. This may prove useful for target samples containing very low concentrations of these particular transcripts. Nevertheless, because our goal was to utilize modifications that would increase hybridization signals for the vast majority of target-probe duplexes, we subsequently focused our efforts on the samples containing DAP.

We next determined whether the increases in signal intensity observed with DAP substitution were similar in magnitude to those observed when the hybridization time is increased. We therefore hybridized two micrograms of cRNA with no modifications or two micrograms with DAP at a 1:1 molar ratio to adenosine for 18 hours and hybridized two micrograms of cRNA with no modifications for 42 hours. We generated two scatterplot and overplayed these scatterplot on each other (Figure 3A). The first scatterplot compares the intensities of the 18 hour control with those of the 42 hour control (orange signals). The second scatterplot compares the intensities of the 18 hour control with those of the DAP-modified cRNA (blue signals). The longer hybridization time increases the median signal intensity by 47% +/- 2%. The increase is more pro-

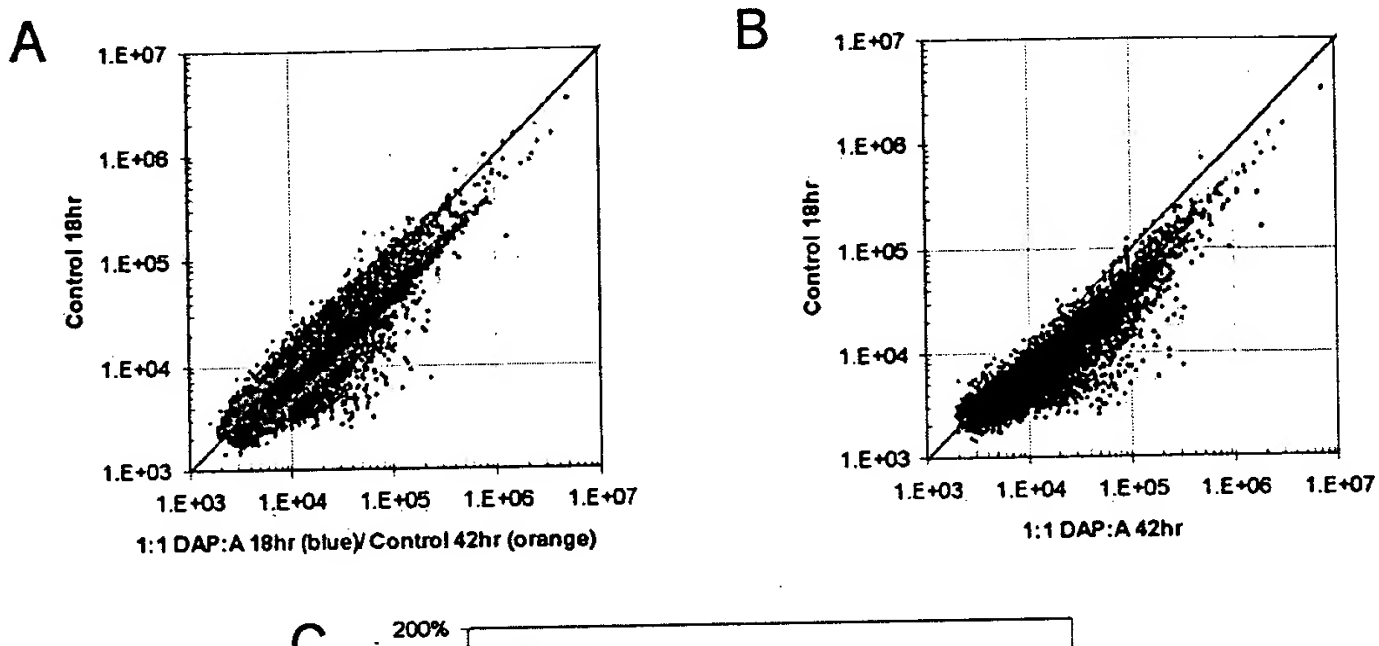


Figure 3
Effects of DAP substitution are similar, and additive, to those seen with an increased hybridization time. (A) Comparison of intensities obtained with unmodified cRNA hybridized for 18 hours versus partially DAP-modified cRNA for 18 hours (blue data points) transposed on a comparison of intensities obtained with unmodified cRNA hybridized for 18 hours versus unmodified cRNA hybridized for 42 hours (orange data points). (B) Comparison of intensities obtained with unmodified cRNA hybridized for 18 hours versus partially DAP-modified cRNA hybridized for 42 hours. (C) Plot showing the relative increases with respect to the 18 hour, unmodified control for increased hybridization time (18 hour ⇒ 42 hour), 1:1 DAP:A substitution, and a combination of increased hybridization time/DAP substitution. The total number of probes was divided into three equally sized bins according to their signal intensities, with the bins representing low, medium, and high expressers. Relative percent increases for each probe were derived from the equation: 100 * (modified sample signal [time, DAP, or combination of both] - control sample signal) / control sample signal, and are shown with standard deviation error bars.

nounced for the medium and high signals. As shown before, the DAP-modified cRNA generated hybridization signals which were also increased relative to the 18 hour control hybridization. However, the increase seemed more pronounced for the low and medium signals. We conclude that both DAP modification and an increased hybridization time can affect the hybridization reaction but in very different ways.

Because of the differential effects of increased hybridization time and DAP substitution on low and high expressers, we wanted determine whether the increases seen of using both DAP modification and an extended hybridization time could be additive. We therefore hybridized unmodified (control) cRNA for 18 hours and a DAP-modified cRNA for 42 hours (Figure 3B). The median slide intensity increased by 110% +/- 17%, which is approximately double the increase seen with either the 1:1 DAP:A or the longer hybridization time by themselves. When all of the approximately 9,000 probes are divided into three equally sized bins representing low, medium, and high expressers according to their respective signal in-

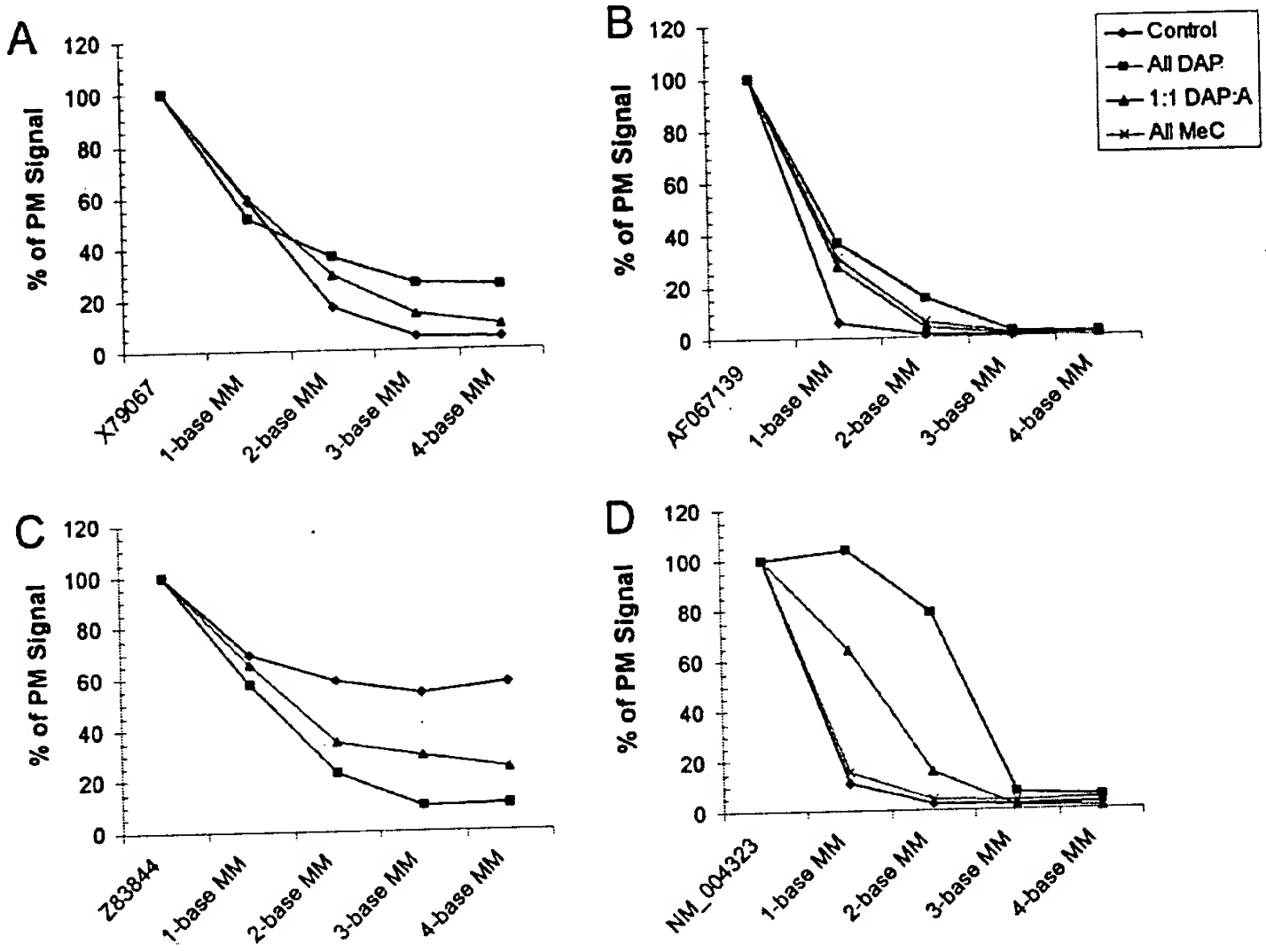


Figure 4
Mismatch discrimination of DAP- or MeC-substituted and control samples. Specificity was determined by introducing single or multiple adjacent, centrally located mismatches in several test probes. Signal intensities of mismatch (MM) probes are plotted as a function of the percent of the perfect match (PM) control for unmodified (diamonds), partially DAP-modified (triangles), fully DAP-substituted (boxes), and fully-MeC-substituted (x) cRNA targets for the transcripts corresponding to the following accession numbers: (A) (X79067), (B) (AF067139), (C) (Z83844), and (D) (NM_004323).

tensities, the effects of increased hybridization time and DAP become very evident (Figure 3C). Figure 3C shows that the relative increases in signal intensities due to increased hybridization time are biased towards the medium and high expressers, with signals increasing an average of 48% and 84%, respectively, whereas increases due to DAP substitution are more pronounced for low and medium expressers (57% and 81% increase, respectively). Together, the two modifications of increased hybridization time and DAP substitution can be used in concert to amplify signal intensities for the entire range of probe signals, and as figure 3C indicates, these boosts in signals are approximately additive for the 42 hour, DAP-containing samples.

We next determined the effect of modified NTPs on hybridization specificity (the ability to distinguish sequences up to a certain homology). We addressed this issue by designing probes which had one to four adjacent, centrally located mismatches and comparing hybridization signals generated from these mismatched probes to signals generated from their corresponding perfect matches. In the control situation (hybridization of unmodified cRNA), the hybridization intensity decreased as the number of mismatches increased, with two mismatches generally destabilizing the duplex sufficiently to reduce the hybridization signal to 0% of the parent signal. There was one exception (Figure 4C) where, even in the presence of four mismatches, the signal was not reduced below 60% of the

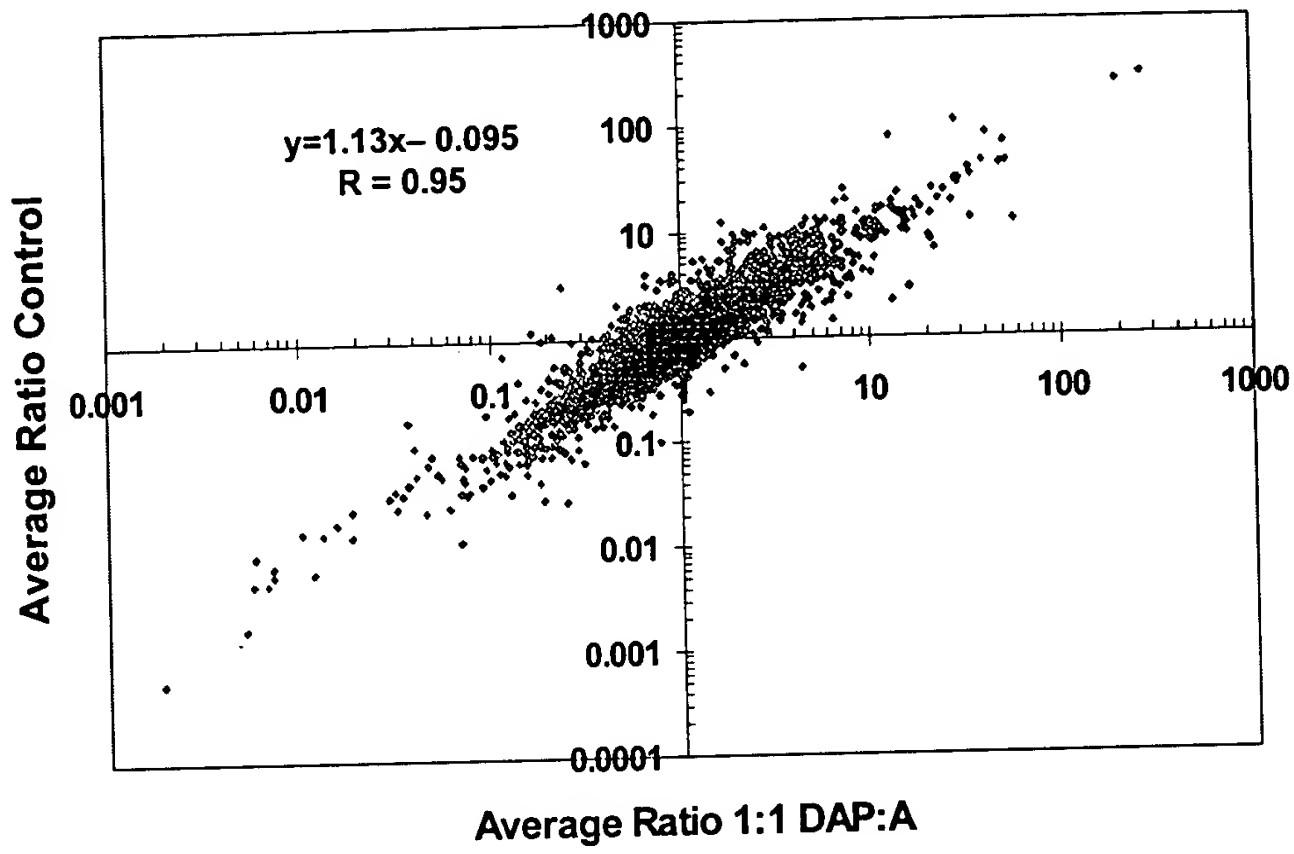


Figure 5
Correlation of kidney to lymphoma differential expression ratios using partially DAP-modified and unmodified control samples. Duplicate hybridizations were performed for each sample, resulting in four possible calls for kidney to lymphoma ratios for each probe (i.e., K1:L1, K1:L2, K2:L1, K2:L2). Data plotted are the average kidney to liver ratio for the four calls on both axes.

parent signal. This signal was low in the perfect match and, therefore, this situation most likely reflects a lowering to the noise levels. In fact, the signal was lowered to below threshold levels in the presence of mismatches. When equimolar ratios of DAP and unmodified ATP were incorporated into the cRNA target (1:1 DAP:A), specificity was not significantly affected in two of the four probe sequences (Figure 4A and 4B). However, in one probe sequence (Figure 4C), specificity was enhanced. For this test probe, the unmodified sample produced unusually high signals for base mismatches. Nevertheless, other groups have seen similar improvements in mismatch discrimination with diaminopurine-containing oligo mers for different applications [13]. A fourth probe sequence demonstrated a smaller destabilizing effect of a single mismatch with the partially modified cRNA compared to the control (Figure 4D). In this sequence, the hybridization was reduced to the same extent with three mismatches using either the control or partially modified cRNA. Full substitution of adenosine by DAP showed varied effects on specificity in the four different probe sequences. In two of the probe sequences, the fully modified cRNA behaved similarly to the control cRNA (Figures 4A and 4B). In a third probe sequence, the specificity improved for the fully modified cRNA, as it did for the partially modified cRNA (Figure 4C). The improvement in specificity for this probe depended on the extent of DAP modification. In a fourth probe sequence, the fully modified cRNA showed a dramatic decrease in specificity with one- or two-base mismatches, although three mismatches reduced the hybridization intensity to ~10% of that of the perfect match. The loss in specificity for this probe sequence also depended on the extent of DAP modification, with one, two, or three mismatches required to reduce the hybridization intensity to ~ 10% for the control, partially modified, and fully modified, respectively, cRNA. A fifth test probe showed hybridization signal intensity below the negative control threshold in both the control and modified ATP samples, suggesting either an absence of the transcript from the target sample or abundance too low to quantify (data not shown). Repeat experiments with cRNA containing modified ATP supported the general specificity trend: unmodified ATP > 1:1 DAP:A > all DAP (data not shown). Attempts to increase specificity in the all DAP condition by raising the temperature during hybridization from 37°C to 42°C or 47°C were unsuccessful (data not shown). Others have been able to distinguish sequences of up to 90% homology [6] using the same plat-

form (3-base mismatch/ 30-mer oligonucleotide signal near or below negative control threshold), and we support their findings here with the control and 1:1 DAP:A condition. The fully MeC-substituted cRNA targets were examined with respect to specificity in two test sequences (Figures 4B and 4D). In one test sequence (Figure B), the fully MeC-substituted cRNA targets behaved nearly identically to the partially DAP-substituted targets. In the other test sequence (Figure 4D), the MeC-substituted cRNA behaved nearly identically to the unmodified cRNA target and showed better specificity than that of the partially DAP-substituted cRNA target with respect to the effect of one mismatch. We conclude that the specificity of fully MeC-substituted cRNA and partially DAP-substituted cRNA targets are, for the most part, equivalent and essentially no loss in specificity (as determined by the effect of two mismatches) is observed with these modified targets.

An additional metric of specificity on the Codelink bioarrays is the use of 54 negative control bacterial probes spotted in 4X redundancy, which were designed and empirically shown not to cross hybridize to human transcripts [6]. While the all DAP condition resulted in an increase in the hybridization intensity for three of the negative control probes, the 1:1 DAP:A condition produced lower background signals similar to those of the unmodified control (data not shown). Thus, increases in hybridization signal intensities in the 1:1 DAP:A condition are attributable to specific modified cRNA target/ DNA probe interactions. We conclude that partial substitution of adenosine with DAP does not significantly compromise specificity and was, therefore, investigated further.

Effect of DAP incorporation on differential expression ratios

Ultimately, the goal of any global expression profiling system measuring relative transcript abundance is to accurately and reproducibly determine the changes in expression levels between different target samples. We therefore determined if the increases in signal intensity caused by DAP incorporation had an effect on differential ratio calls. Kidney and lymphoma cRNA samples containing either no modified NTPs or 1:1 DAP:A were hybridized in duplicate to Human Uniset I arrays; average kidney to lymphoma ratios were calculated and plotted (Figure 5) after removing signal outliers defined as a two-fold difference between replicates. Differential expression ratios demonstrate very good correlation between the unmodified control sample and the 1:1 DAP:A sample ($r = 0.95$) on the Human Uniset I microarray.

In addition to investigating the correlation of the differential expression ratios generated from modified and unmodified targets, we determined whether DAP incorporation affected the variability of the ratio calls. One method to measure microarray reproducibility is by calculating the coefficients of variation (CVs) for each replicate probe across arrays. Likewise, CVs can also be calculated for multiple ratio calculations across different tissue samples. Figure 6 shows a plot of the CVs of the ratios as a function of the mean ratios for all of the data points, *regardless of the intensity level of the probe.* Consistent with earlier findings [6], the majority of the CVs in the of the kidney to lymphoma ratios are below 30%, with an average of 13.2%, and the variability increases as the ratio approaches unity for the unmodified targets (Figure 6A). Figure 6B shows that the 1:1 DAP:A condition produces ratios with similar overall variability (average CV of 12.4%) compared to the unmodified control targets. When the CVs in the ratios generated using either the control or partially modified cRNA are binned according to the magnitude of the CV, the partially modified cRNA generates more CVs in the less than 20% bins (Figure 6C). These data demonstrate that the low variability in ratio calls that we are able to routinely obtain [6] are maintained during DAP incorporation. However, a closer inspection of the ratios in Figures 6A and 6B reveals that a possible caveat of DAP use is a slight compression of ratios along the entire range of calls. This compression is also observed when the observed ratio is binned according to the magnitude of the fold change (Figure 6D). The partially modified cRNA generates more ratios in the one to 1.5 fold change bin and fewer ratio changes in the other bins. This compression is not significant but also manifests itself in the slope of the correlation line in Figure 5. The slope of this line was found to be greater than unity.

The differential expression ratios calculated when using DAP (Figures 5 and 6) suggest that incorporation of DAP has little effect on correlation and reproducibility of ratio calls between tissues. However, the ability to accurately discriminate between signals due to true hybridization events and what may be considered background noise is paramount to making accurate differential assessments. A lower limit of detection was previously defined using our platform by developing a negative control threshold in order to assign a confidence level to such signal or noise queries [6]. Briefly, this threshold was determined by taking the mean signal of bacterial negative control probes mentioned above (minus a 10% trim to account for weak cross hybridization of high expressers or for true hybridization to sequences not in the database) and adding three standard deviations (99.7% confidence). Using only probes that were at or above the negative control threshold in both tissues, we observed a significantly greater number of probes in the 1:1 DAP:A sample for which we could confidently assign a ratio (Figure 7). As Figure 7 shows, samples containing DAP generated over 400 more above-threshold ratios than the control samples contain-

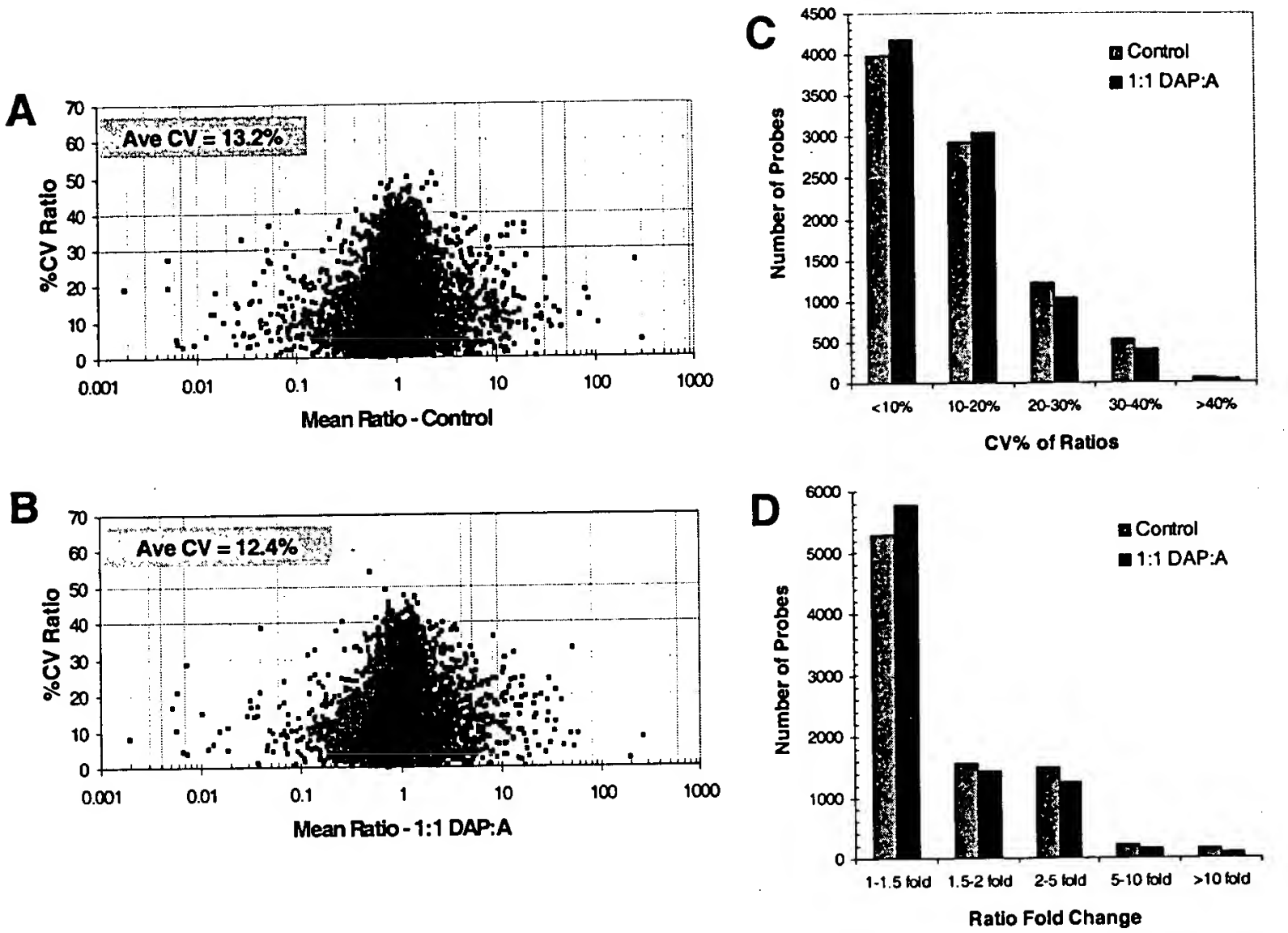


Figure 6
Effect of partial DAP substitution on the variability and magnitude of differential expression ratios. (A and B) CVs of the ratios plotted as a function of the mean ratio for unmodified (A) and modified (B) samples. Signal outliers, defined as any pair of signals which were greater than two-fold different in intensity, were removed from all samples. (C) Plot showing the number of probes having a designated CV across a range of CVs for control and partially DAP-modified cRNA targets. (D) Plot showing the number of probes having a designated fold change (ratio) across a range of ratios for control and partially DAP-modified cRNA targets.

ing only unmodified NTPs. Most of these ratios identified only in the 1:1 DAP:A sample were small in magnitude, with fold changes between 1.1 and 1.9.

Discussion

The effect of modified NTPs on hybridization intensities

Two important issues for identifying differentially expressed genes using DNA microarrays are 1) methods to increase signal intensity and 2) the ability to accurately separate true hybridization signals from background noise at the low end of the signal dynamic range. In this study, we utilized modified nucleoside triphosphates with the aim of enhancing hybridization signal intensity across a wide range of probe sequences. Modified ribo- and deoxyribonucleoside analogs have been used in a variety of applications [12–14], but their use has been limited in the field of microarrays [23]. We demonstrate that 2-aminoadenosine or 2,6-diaminopurine (DAP), an analog of adenosine, significantly increases signal intensity for a wide range of probe sequences, whereas C-5 methylated pyrimidine analogs of cytosine and uridine decreased signal intensities across the entire range of probe sequences on our platform (Figure 2).

Although incorporation of DAP did not uniformly increase the hybridization stability for all target-probe duplexes, even moderate incorporation of DAP (1:1 DAP:A) resulted in signal intensity increases (up to 30-fold) for the majority of probes on our platform. More importantly, a significant number of probes that went previously

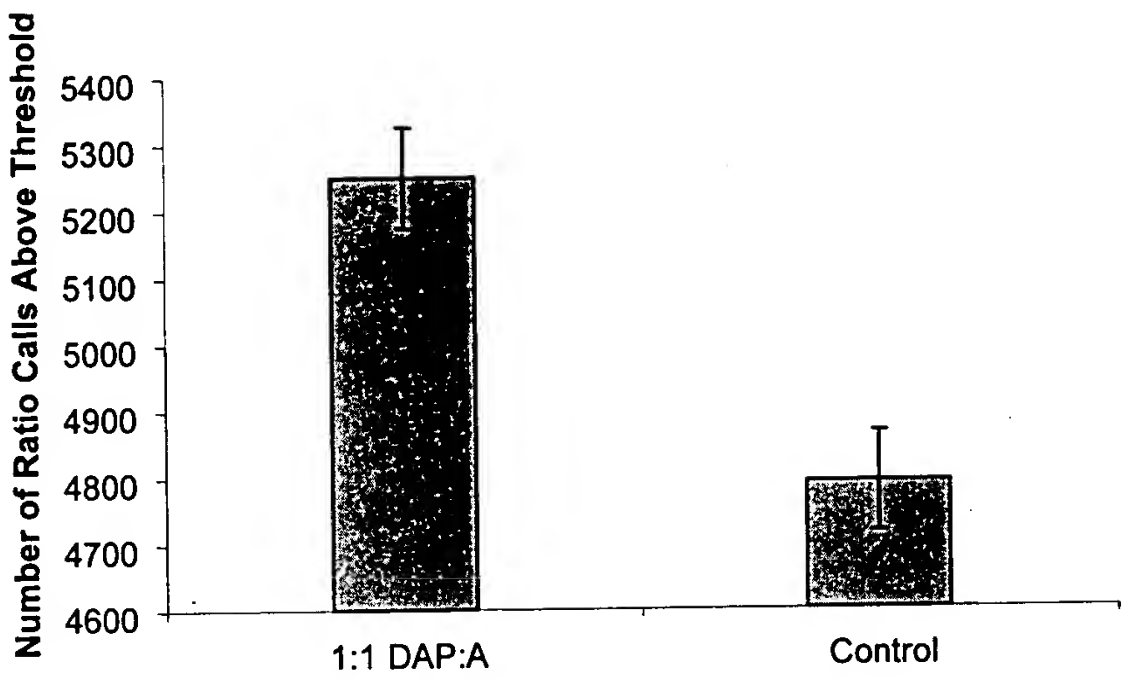


Figure 7
Number of above-threshold ratios generated by partially DAP-modified and unmodified target samples. A negative control threshold was imposed to define the lower limit of detection and was calculated by taking the 80% trimmed mean (top 10% and bottom 10% of signals removed from the population) of 216 negative control probes and adding three standard deviations to the mean (99.7% confidence). Embryonic kidney and Burkitt's lymphoma samples were performed in duplicate and data were screened for probes that displayed above-threshold signal intensities. Kidney to lymphoma ratio calls were calculated using only above-threshold probe signals with a maximum of four ratios for each probe. The data represent the average of the number of calls for any two kidney and lymphoma slides +/- standard deviation error bars.

undetected (i.e., below a negative control threshold; [6]) in the control sample exhibited signal intensities above this threshold in DAP-modified samples. Utilization of such a modification may prove especially valuable for low abundant transcripts which may be difficult to quantify because of weak hybridization kinetics. As shown in Figures 2B and 2C, the largest increases in signal intensities occurred for low and medium expressers relative to the unmodified control sample. It is apparent that the effect of DAP depends not only on the sequence compositions of target and probe, but also on the absolute abundance of specific transcripts in the target sample.

It is of interest to note that the results we report here differ from those obtained by others [23]. Those studies reported increases in signal intensity for targets containing 5-methyl UTP while completely substituted targets containing DAP had the opposite or no effect. Although we cannot completely account for the disparity of the results, one key difference in methodology between the studies is that their group was not able to normalize concentrations for modified and unmodified ATP targets because of dramatic differences in cRNA yield. Therefore, mass input of cRNA target samples were much greater for unmodified samples compared to the DAP-containing samples. Because the sample preparation used in this protocol is highly robust and reproducible [24], we were able to use equivalent mass inputs of all modified and unmodified targets during hybridization, allowing us to present our results quantitatively rather than qualitatively. However, we agree that complete substitution of DAP for adenosine may cause specificity to decline in certain sequence contexts, which may limit its usefulness in microarray applications. However, we have found that reducing the proportion of DAP incorporated into target samples (e.g., incorporating a 1:1 ratio of DAP:adenosine) is one approach to avoid such decreases in specificity. Other factors that may contribute to the dissimilarity of the results include differences in probe sequence and size, microarray fabrication/chemistries, labeling/detection methods, and the vigorous mixing that is employed in our microarray experiments. We have previously shown the dramatic impact of mixing on hybridization signal intensities [6]. Presumably, this impact is due to the effect of three dimensional diffusion of the target molecules and to the effect on how many probes reach equilibrium in the hybridization. It is possible that the degree of mixing during the hybridization can affect the number of probes which show a difference in intensity

with substitution of DAP for adenosine as well as the magnitude of this difference. Given these differences, it is not surprising that our results are not in line with those previously reported.

The results in our study are in good agreement with previous reports which include a wide variety of interaction types, including DNA:DNA, DNA:RNA, and DNA:PNA interactions. [11–13,25]. These studies have shown that DAP substitution in oligonucleotide probes can increase thermodynamic affinities as reported directly by increases in T_m and indirectly through increases in hybridization intensities.

Biophysical considerations in using modified NTPs

Modified nucleotides appear to drive the equilibrium (particularly for low copy transcripts binding with their targets) towards duplex formation. These findings suggest that DAP use could be exploited in RNA samples that are limited, such as tumor biopsies. Based on our experiments with extended hybridization times (Figure 3), use of DAP may prove to be a useful alternative to increase array sensitivity for studies that are unable to generate enough target to hybridize or are unable to increase hybridization times because of throughput constraints.

Why do we see differential effects of DAP, MeC, and MeU on hybridization intensities? The fact that DAP was able to increase hybridization intensities while MeC and MeU decreased hybridization intensities (Figures 2D and 2E) is interesting in light of the fact that DAP is a purine incorporated into an RNA strand while MeC and MeU are pyrimidines incorporated into an RNA strand. Many studies have shown that a ribopurine rich strand bound to a deoxypyrimidine rich strand has a higher stability than the corresponding all deoxy strands which, in turn, have a higher stability than a deoxypurine rich strand bound to a ribopyrimidine rich strand [26,27]. These differences also suggest that further studies examining DAP or MeC incorporation into first strand cDNA, followed by formation of a DNA-DNA duplex, may not necessarily show the same effects as incorporation of these modified nucleotides into cRNA. Finally, earlier studies have noted that methylation of cytosines (generating MeC) in ribopolynucleotides stabilizes duplexes to the same extent as substitution of thymine (MeU) for uracil [28]. Thus, although it is still surprising that MeC and MeU decrease hybridization intensities, it is not surprising that the magnitude of the effects generated by MeC and MeU residues are similar.

Could the differential effects of DAP versus MeC or MeU or the differential effects of DAP on different probe sequences also be related to structural consequences? Earlier studies have shown that DAP, in DNA fragments, can widen the minor groove, as detected by reactivity towards uranyl nitrate or susceptibility to Dnase cleavage [29], and can relieve compressions in the minor groove associated with A-tracts [30]. In contrast, NMR and molecular modeling studies of an oligonucleotide duplex containing MeC have demonstrated that, while this duplex still has a B-DNA conformation, there are differences in the structural parameters and thermal stability relative to the unmodified duplex [31]. An earlier study also found, in two closely related octanucleotide duplexes, that although the methylated duplexes retained their B-DNA conformation, different structural and thermal stability effects were seen [32]. Although these studies have been carried out with DNA-DNA and not DNA-RNA duplexes, it is possible that MeC substitution may have different structural consequences than DAP substitution.

We have shown that while complete substitution of DAP for ATP may cause decreased base mismatch discrimination in certain sequence contexts, equimolar ratios of adenosine and DAP in cRNA target samples did not have significant detrimental effects on specificity (Figure 4). Remarkably, in one sequence context, the DAP modification enhanced specificity against mismatches. Enhanced specificity against mismatches has also been observed by other groups when DAP is incorporated into PNA oligomers [13] and after C5-(1-propynyl)ation of pyrimidines in DNA-RNA duplexes [33].

The effect of modified NTPs on differential expression ratios

The use of DAP to increase hybridization signals is supported by the results that show equivalency of both ratio calls and ratio variability in 1:1 DAP:A and control samples (Figures 5 and 6). When all data points are considered in the differential expression ratio analyses, the two methods appear equivalent. However, when only those data points which are above a threshold value [6] are used, ensuring a higher confidence that these intensities represent true expression levels and not noise or weak cross hybridization, the DAP-modified cRNA enabled an ~ 10% increase in the number of ratios that could be generated (Figure 7). Such increases are critical in order to sample as many genes as possible in any given microarray experiment. For example, some investigators have analyzed only data for which the signal intensity is greater than approximately 0.4% of the total signal range in both channels [34]. Although this method minimizes the variance associated with the ratios, many genes and their expression changes are missed. Other platforms use a minimal intensity level to determine whether a gene is present or absent, prior to including this gene in expression analyses [35]. It is plausible that DAP modification, by increasing intensity levels, particularly for low expressors, could enable wider coverage in genome-wide expression analyses.

In summary, incorporation of modified nucleotides may not only lead to discovery and better quantitation of rare expressers in complex samples but also impact other aspects of the microarray experiment such as probe design prior to array fabrication and the statistical design of microarray experiments. The target preparation procedure (generation of amplified cRNA or of first strand cDNA) and the subsequent duplex formation may dictate the choice of which modified nucleotide to use, as we and others have found differential effects with various nucleotides. We believe further investigation in this area could be fruitful.

Methods

Target preparation

Five µgs of human embryonic kidney or Burkitt's lymphoma total RNA (Ambion, Inc., Austin, TX) were added to a reaction mix in a final volume of 12 µl, containing 0.5 pmol T7-$(dT)_{24}$ oligonucleotide primer. The mixture was incubated for 10 minutes at 70°C and chilled on ice. With the mixture remaining on ice, 4 µl of 5X first-strand buffer, 2 µl 0.1 M DTT, 1 µl of 10 mM dNTP mix and 1 µl Superscript™ II RNaseH$^-$ reverse transcriptase (200 U/µl) was added to make a final volume of 20 µl, and the mixture incubated for one hour in a 42°C water bath. Second-strand cDNA was synthesized in a final volume of 150 µl, in a mixture containing 30 µl of 5X second-strand buffer, 3 µl of 10 mM dNTP mix, 4 µl of *E. coli* DNA polymerase I (10 U/µl) and 1 µl of RNase H (2 U/µl) for 2 hours at 16°C. The cDNA was purified using a Qiagen QIAquick purification kit, dried down, and resuspended in IVT reaction mix, containing 3.0 µl nuclease-free water, 4.0 µl 10X reaction buffer, 4.0 µl 75 mM ATP, 4.0 µl 75 mM GTP 3.0 µl 75 mM CTP, 3.0 µl 75 mM UTP, 7.5 µl 10 mM Biotin-11-CTP, 7.5 µl 10 mM Biotin 11-UTP and 4.0 µl enzyme mix (unmodified control condition). Commercially available 2-aminoadenosine-5'-triphosphate, 5-methylcytidine-5'-triphosphate, and 5-methyluridine-5'-triphosphate (TriLink Biotechnologies, Inc., San Diego, CA) were substituted for ATP, CTP, and UTP, respectively, in separate reactions containing either complete substitution, 1:1, or 1:3 ratio of modified: unmodified NTP, keeping molar input of nucleotide constant. The reaction mix was incubated for 14 hours at 37°C and cRNA target purified using an RNeasy® Kit (Qiagen). cRNA yield was quantitated by measuring the UV absorbance at 260 nm, and fragmented in 40 mM Tris-acetate (TrisOAc), pH 7.9, 100 mM KOAc, and 31.5 mM MgOAc, at 94°C, for 20 minutes. This typically resulted in fragmented target with a size range between 100–200 bases.

Array hybridization

Two µg of fragmented target cRNA was used for hybridization of each UniSet Human I Expression Bioarray (Motorola Life Sciences) containing 9589 probes (representing 9,203 unique accession numbers (genes), corresponding to approximately 8,935 unique clusters and 386 control probes, selected initially from GenBank Unigene build #125) or for hybridization to a microarray containing probes corresponding to 1100 human genes, each spotted 6 times per array. All probes on these microarrays are 30-mer oligonucleotides spotted by piezoelectric technologies and covalently attached to a polymeric matrix [6]. These microarrays were hybridized, washed, and processed using a direct detection method of the biotin-containing transcripts by a Streptavidin-Alexa647 conjugate as previously described [6]. Processed slides were scanned using an Axon GenePix Scanner with the laser set to 635 nm, a PMT voltage of 600, and a scan resolution of 10 microns.

Data analysis

Slides were scanned and images for each slide were quantitated using CodeLink Scanning and Analysis Software (Motorola Life Sciences). Signal intensities for each spot were calculated by summation of the pixel intensities for each spot, followed by local background subtraction (based on the median pixel intensity of the area surrounding each spot). Whole array data normalization, when used, was performed independently for each slide by dividing each spot's intensity (after background subtraction) by the median signal intensity of all test probes. All false positives, determined by visual inspection of the images, which were greater than 2-fold different between duplicate arrays were removed.

Digestion and chromatography of cRNA

Four units of P1 nuclease were used to digest 20–50 µg of cRNA to generate nucleotide monophosphates. The enzyme was incubated with the cRNA at 55°C for 6 hours, then for 6 hours at 37°C in the presence of 10 units of calf intestine alkaline phosphatase to generate the nucleosides. The digested products were purified using Microcon YM-3 columns followed by centrifugation at 8000 *g* for 30–60 minutes. The mix was then concentrated using a SpeedVac to 100 µl. This solution was analyzed on an HPLC column equilibrated with 0.03 M TEAA (Solvent A) at a flow rate of 1 ml/min. The following gradient was used: 0–1% Solvent B (95% AcCN, 5% Solvent A1) over 5 minutes, 1–15% Solvent B over the next 15 minutes, 15–45% Solvent B over the next 30 minutes, 45–100% Solvent B over the next 20 minutes, and hold at 100% B for 2 minutes. The concentration of the heterocycles was determined by the absorbance values at 260 nm (the wavelength where maximal absorption occurs for the heterocycles) and the biotin-containing nucleoside concentrations were determined by the absorbance values at 294 nm (the wavelength where maximal absorption occurs for biotinylated cytosine and biotinylated uridine). The biotin-11-UTP peak was measured at an absorbance

of 294 nm with the extinction coefficient = 13000 $M^{-1}cm^{-1}$. The 2-amino ATP was measured at an absorbance of 260 nm with the extinction coefficient = 9894 $M^{-1}cm^{-1}$.

Authors' contributions

AN performed the target preparations, microarray hybridizations, and data analyses, and drafted the manuscript. CZ performed the digestion and chromatography of the cRNAs. DD participated in design of the study and data analyses. AM conceived the study, participated in its design and coordination, and aided in drafting the manuscript.

References

1. Heller R, Schena M, Chai A, Shalon D, Bedilion T, Gilmore J, Woolley DE, Davis RW: **Discovery and analysis of inflammatory disease-related genes using cDNA microarrays.** *Proc Natl Acad Sci USA* 1997, **94**:2150-2155
2. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale A-L, Brown PO, Botstein D: **Molecular portraits of human breast tumors.** *Nature* 2000, **406**:747-752
3. Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A, Sampas N, Dougherty E, Wang E, Marincola F, Gooden C, Lueders J, Glatfelter A, Pollock P, Carpten J, Gillanders E, Leja D, Dietrich K, Beaudry C, Berens M, Alberts D, Sondak V, Hayward N, Trent J: **Molecular classification of cutaneous malignant melanoma by gene expression profiling.** *Nature* 2000, **406**:536-540
4. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: **Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.** *Science* 1999, **286**:531-537
5. Hu GK, Madore SJ, Moldover B, Jatkoe T, Balaban D, Thomas J, Wang Y: **Predicting splice variant from DNA chip expression data.** *Genome Res* 2001, **11**:1237-1245
6. Ramakrishnan R, Dorris D, Lublinsky A, Nguyen A, Domanus M, Prokhorova A, Gieser L, Touma E, Lockner R, Tata M, Zhu X, Patterson M, Shippy R, Sendera T, Mazumder A: **An assessment of Motorola Codelink™ microarray performance for gene expression profiling applications.** *Nucleic Acids Res* 2002, **30**:e30
7. Stears RL, Getts RC, Gullans SR: **A novel, sensitive detection system for high-density microarrays using dendrimer technology.** *Physiol Genomics* 2000, **9**:93-99
8. Kane MD, Jatkoe TA, Stumpf CR, Lu J, Thomas JD, Madore SJ: **Assessment of the sensitivity and specificity of oligonucleotide (50 mer) microarrays.** *Nucleic Acids Res* 2000, **28**:4552-4557
9. Weiler J, Gausepohl H, Hauser N, Jensen ON, Hoheisel JD: **Hybridisation based DNA screening on peptide nucleic acid (PNA) oligomer arrays.** *Nucleic Acids Res* 1997, **25**:2792-2799
10. Broude NE, Woodward K, Cavallo R, Cantor DR, Englert D: **DNA microarrays with stem-loop DNA probes: preparation and applications.** *Nucleic Acids Res* 2001, **29**:e92
11. Cheong C, Tinoco I, Chollet A: **Thermodynamic studies of base pairing involving 2,6-diaminopurine.** *Nucleic Acids Res* 1988, **16**:5115-5122
12. Matray T, Gamsey S, Pongracz K, Gryaznov S: **A remarkable stabilization of complexes formed by 2,6-diaminopurine oligonucleotide N3' → P5' phophoramidates.** *Nucleosides Nucleotides Nucleic Acids* 2000, **19**:1553-1567
13. Haaima G, Hansen H, Christensen L, Dahl O, Nielsen P: **Increased DNA binding and sequence discrimination of PNA oligomers containing 2,6-diaminopurine.** *Nucleic Acids Res* 1997, **25**:4639-4643
14. Prosnyak MI, Veselovskaya SI, Myasnikov VA, Efremova EJ, Potapov VK, Limborska SA, Sverdlov ED: **Substitution of 2-aminoadenine and 5-methylcytosine for adenine and cytosine in hybridization probes increases the sensitivity of DNA fingerprinting.** *Genomics* 1994, **21**:490-494
15. Hoheisel JD, Craig AG, Lehrach H: **Effect of 5-bromo- and 5-methyldeoxycytosine on duplex stability and discrimination of the NotI octadeoxynucleotide. Quantitative measurements using thin-layer chromatography.** *J Biol Chem* 1990, **265**:16656-16660
16. Khudiakov IIa, Kirnos MD, Aleksandrushkina NI, Vaniushin BF: **[2,6-Diaminopurine--a new adenine substituent base in the DNA of cyanophage S-2].** *Dokl Akad Nauk SSSR* 1977, **232**:965-968
17. Rackwitz HR, Scheit KH: **The stereochemical basis of template function.** *Eur J Biochem* 1977, **72**:191-200
18. Vaish NK, Fraley AW, Szostak JW, McLaughlin LW: **Expanding the structural and functional diversity of RNA: analog uridine triphosphates as candidates for in vitro selection of nucleic acids.** *Nucleic Acids Res* 2000, **28**:3316-3322
19. Hacia JG, Novotny EA, Mayer RA, Woski SA, Ashlock MA, Collins FS: **Design of modified oligodeoxynucleotide probes to detect telomere repeat sequences in FISH assays.** *Nucleic Acids Res* 1999, **27**:4034-4039
20. Stump MD, Cherry JL, Weiss RB: **The use of modified primers to eliminate cycle sequencing artifacts.** *Nucleic Acids Res* 1999, **27**:4642-4648
21. Fidanza JA, McGall GH: **High-density nucleoside probe arrays for enhanced hybridization.** *Nucleosides Nucleotides* 1999, **18**:1293-1295
22. Timofeev E, Mirzabekov A: **Binding specificity and stability of duplexes formed by modified oligonucleotides with a 4096-hexanucleotide microarray.** *Nucleic Acids Res* 2001, **29**:2626-2634
23. Hacia JG, Woski SA, Fidanza J, Edgemon K, Hunt N, McGall G, Fodor SP, Collins FS: **Enhanced high density oligonucleotide array-based sequence analysis using modified nucleoside triphosphates.** *Nucleic Acids Res* 1998, **26**:4975-4982
24. Dorris DR, Ramakrishnan R, Trakas D, Dudzik F, Belval R, Zhao C, Nguyen A, Domanus M, Mazumder A: **A highly reproducible, linear, and automated sample preparation method for DNA microarrays.** *Genome Res* 2002, **12**:976-984
25. Lebedev Y, Akopyants N, Azhikina T, Shevchenko Y, Potapov V, Stecenko D, Berg D, Sverdlov E: **Oligonucleotides containing 2-aminoadenine and 5-methylcytosine are more effective as primers for PCR amplification than their nonmodified counterparts.** *Genet Analy* 1996, **13**:15-21
26. Lesnik EA, Freier SM: **Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure.** *Biochemistry* 1995, **34**:10807-10815
27. Gyi JI, Conn GL, Lane AN, Brown T: **Comparison of the thermodynamic stabilities and solution conformations of DNA/RNA hybrids containing purine-rich and pyrimidine-rich strands with DNA and RNA duplexes.** *Biochemistry* 1996, **35**:12538-12548
28. Collins M, Myers RM: **Alterations in DNA helix stability due to base modifications can be evaluated using denaturing gradient gel electrophoresis.** *J Mol Biol* 1987, **198**:737-744
29. Bailly C, Mollegaard NE, Nielsen PE, Waring MJ: **The influence of the 2-amino group of guanine on DNA conformation. Uranyl and Dnase I probing of inosine/diaminopurine substituted DNA.** *EMBO Journal* 1995, **14**:2121-2131
30. Lavigne M, Buc H: **Compression of the DNA minor groove is responsible for termination of DNA synthesis by HIV-1 reverse transcriptase.** *J Mol Biol* 1999, **285**:977-995
31. Marcourt L, Cordier C, Couesnon T, Dodin G: **Impact of C5-cytosine methylation on the solution structure of d(GAAAACGTTTTC)2. An NMR and molecular modeling investigation.** *Eur J Biochem* 1999, **265**:1032-1042
32. Lefebvre A, Mauffret O, el Antri S, Monnot M, Lescot E, Fermandjian S: **Sequence-dependent effects of CpG cytosine methylation. A joint 1H-NMR and 31P-NMR study.** *Eur J Biochem* 1995, **229**:445-454
33. Barnes III TW, Turner DH: **C5-(1-propynyl)-2'-deoxy-pyrimidines enhance mismatch penalties of DNA:RNA duplex formation.** *Biochemistry* 2001, **40**:12738-12745
34. Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, Iyer V, Jeffrey SS, Van de Rijn M, Waltham M, Pergamenschikov A, Lee JCF, Lashkari D, Shalon D, Myers TG, Weinstein JN, Botstein D, Brown PO: **Systematic variation in gene expression patterns in human cancer cell lines.** *Nature Genetics* 2000, **24**:227-235

35. Hill AA, Hunter CP, Tsung BT, Tucker-Kellogg G, Brown EL: **Genomic analysis of gene expression in C. elegans.** *Science* 2000, **290**:809-812

Nucleic Acids Research, 1997, Vol. 25, No. 15 3059–3065

Modification of DNA duplexes to smooth their thermal stability independently of their base content for DNA sequencing by hybridization

Hong-Khanh Nguyen, Pascal Auffray+, Ulysse Asseline, Daniel Dupret[1] and Nguyen Thanh Thuong*

Centre de Biophysique Moléculaire, CNRS, rue Charles Sadron, 45071 Orléans Cedex 02, France and [1]Appligène-Oncor, Parc d'Innovation, BP 72, 67407 Illkirch, France

Received April 14, 1997; Revised and Accepted June 9, 1997

ABSTRACT

The possibility of equalizing DNA duplex stability is essential for the application of sequencing by hybridization. In this paper we describe a new strategy to obtain DNA duplexes with a thermal stability independent of their base content. Modified *C bases have been developed and incorporated into oligonucleotides. The influence of these modifications on duplex stability has been studied by absorption spectroscopy, thus allowing selection of *N*-4-ethyl-2′-deoxycytidine ($d^{4Et}C$), which hybridizes specifically with natural dG to give a $G^{4Et}C$ base pair whose stability is very close to that of natural AT base pairs. Duplexes built with AT and/or $G^{4Et}C$ base pairs exhibit thermal stabilities independent of their base content in a classical buffer solution, thus enabling control of the stability of DNA hybrids as a function of their length only.

INTRODUCTION

Specific duplex formation between an oligonucleotide and a DNA sequence is the foundation of nucleic acid analysis and enzymatic labeling of DNA fragments. The success of these techniques implies absolute discrimination of perfect hybrids from ones containing mismatches to avoid false positive results and the design of hybrids having equivalent stabilities, which leads to homogeneous results.

The recently reported reverse hybridization technique is based on detection of perfect hybrids formed between one or several labeled DNA fragments with every oligonucleotide of a given length in a complete set immobilized as a two-dimensional matrix (1–3). This method is based on the possibility of discriminating perfect hybrids from those containing mismatches. This can be performed by increasing the temperature to allow dissociation of hybrids with mismatches before those without mismatches. Reverse hybridization could be a method for the design of a fast analysis technique for large DNA sequences. However, the approaches described in the literature using natural oligonucleotides serving as probes have a serious drawback, due to the base composition dependence of duplex stability. It is well known that a GC base pair with three hydrogen bonds is more stable than AT or AU base pairs, which have only two hydrogen bonds. A perfect hybrid built with AT-rich sequences would therefore have a similar or even lower stability than do hybrids built with GC-rich sequences involving one mismatch. This leads to false positive or false negative signals depending on the hybridization temperature and washing conditions.

Several techniques have been studied to alleviate this problem but none of them have been successfully completed. Hybridization studies using TMACl to reduce differential DNA duplex stability according to the base composition have been described and developed by several groups (4–8). However, this process requires multimolar concentrations of TMACl, which is very viscous, leading to manipulation difficulties, and is not adapted to biochemical procedures involving enzymes, such as random priming or LCR, which are carried out under weak ionic concentrations (9). The proposed concentration variation of each particular oligonucleotide (3) or changing of the oligonucleotide length as a function of its base composition would be technically complex and impractical when applied to a large number of oligonucleotide sequences. Another option, the use of modified bases such as 5-ClduU and $2\text{-}NH_{2d}A$ (10), has been examined, but with a great variation in stability. To reduce the stability of GC-rich duplexes, oligonucleotides were built with ribo- and deoxyribonucleotides, but the amount of hybrid obtained greatly decreased as the number of ribo↔deoxyribo transitions increased. Moreover, hybrid thermal dissociation spread over a wide temperature range, making it difficult to discriminate between perfect hybrids and those containing a mismatch (11).

In this paper we describe a new strategy for obtaining DNA duplexes whose thermal stabilities in NaCl solution are independent of their base content. Our approach consists of modification of one of the four natural deoxynucleosides which forms, with the complementary nucleoside, a base pair whose stability is very close to that of the other base pair. To achieve this end, we chose to modify 2′-deoxycytidine (d*C), which hybridizes specifically with natural 2′-deoxyguanosine (dG) to give a G*C base pair having a stability very similar to that of the AT base pair. We describe in this paper the preparation of modified oligonucleotides

*To whom correspondence should be addressed. Tel: +33 2 38 25 55 97; Fax: +33 2 38 63 15 17; Email: asseline@cnrs-orleans.fr

+Present address: Appligène-Oncor, Parc d'Innovation, BP 72, 67407 Illkirch, France

and selection of $d^{4Et}C$, which gives a $G^{4Et}C$ base pair having a stability similar to that of the AT base pair.

MATERIALS AND METHODS

All reagents and solvents were of reagent grade quality and used without further purification. Cytosine arabinoside phosphoramidite was obtained from Glen Research. Unmodified oligonucleotides were from Appligène-Oncor. Absorption spectra were recorded on a UVIKON 860 (Kontron). Absorption studies were carried out on a UVIKON 941 cell changer spectrophotometer (Kontron). Analytical TLC was carried out on Merck 5554 Kieselgel 60F 254 plates and eluted with various eluents: system A, CH_2Cl_2/MeOH (95:5 v/v); system B, CH_2Cl_2/MeOH (90:10 v/v); system C, CH_2Cl_2/AcOEt/Et_3N, (45:45:10 v/v/v). Merck 9385 Kieselgel 60 was used for column chromatography. All 4,4-dimethoxytrityl-containing substances were identified as orange colored spots on TLC plates by spraying with 10% perchloric acid solution. HPLC was performed on a Waters 626 E (system controller) equipped with a Waters 996 photodiode array detector. Analysis and purification by ion exchange chromatography were performed with a FPLC apparatus (Pharmacia). NMR experiments were carried out on a Bruker AM 300 WB spectrometer.

Synthesis of modified nucleosides 1

5′-O-Dimethoxytrityl-N-4-ethyl-2′-deoxycytidine ***1b***. 1,2,4-Triazole (0.047 mol, 3.1 g) was suspended in anhydrous CH_3CN (50 ml) at 0°C followed by addition of $POCl_3$ (0.01 mol, 0.98 ml) with rapid stirring. Triethylamine (0.052 mol, 7.3 ml) was then added dropwise to the slurry stirred at 0°C for an additional 30 min. 5′-*O*-Dimethoxytrityl-3′-*O*-(t-butyldimethylsilyl)-2′-deoxyuridine (0.031 mol, 2 g), obtained as described in the literature (12), dissolved in 6 ml CH_3CN was added dropwise at 0°C. The ice/water bath was removed and the mixure allowed to react with magnetic stirring for 4 h at room temperature. An ethylamine solution (0.15 mol, 16 M in CH_3CN) was added directly to the crude derivative triazolyl obtained above. The reaction was monitored by TLC analysis. The reaction was complete after 2 h. The solution was concentrated under reduced pressure, solubilized in dichloromethane and washed with 5% sodium hydrogen carbonate solution. After being dried over Na_2SO_4 and concentrated to dryness, the residue was purified on a silica gel column using CH_2Cl_2/MeOH/Et_3N (98:1:1 to 97:2:1 v/v/v) as eluent. Yield 81% (1.7 g), R_f system A 0.26, R_f system B 0.7.

The syntheses of 5′-*O*-dimethoxytrityl-3′-*O*-(t-butyldimethylsilyl)-*N*-4-methyl-2′-deoxycytidine, 5′-*O*-dimethoxytrityl-3′-*O*-(t-butyldimethylsilyl)-*N*-4-propyl-2′-deoxycytidine, 5′-*O*-dimethoxytrityl-3′-*O*-(t-butyldimethylsilyl)-*N*-4-allyl-2′-deoxycytidine and 5′-*O*-dimethoxytrityl-3′-*O*-(t-butyldimethylsilyl)-*N*-4-propargyl-2′-deoxycytidine were carried out as described above by replacing ethylamine by methylamine, propylamine, allylamine and propargylamine respectively (13).

5′-*O*-Dimethoxytrityl-3′-*O*-(t-butyldimethylsilyl)-*N*-4-ethyl-2′-deoxycytidine obtained above (0.0025 mol, 1.7 g) was treated with a 1 M tetrabutylammonium fluoride solution (0.005 mol, 5 ml) in THF at room temperature. The reaction, monitored by TLC, was complete after 2 h. The reaction mixture was concentrated under reduced pressure and the residue dissolved in CH_2Cl_2 and washed with 5% sodium hydrogen carbonate solution. After being dried over Na_2SO_4 and concentrated to dryness the obtained residue was purified on a column of silica gel using CH_2Cl_2/MeOH (98:2 to 95:5 v/v v/v) as eluent. Yield **1b** 78% (1.1 g), R_{f1b} system B 0.55. ^1H NMR (DMSO): δ 1.1 (t, 3H, -NH-CH_2-CH_3), 2 (m, 1H, $H_{2''}$), 2.2 (m, 1H, $H_{2'}$), 3.2 (m, 2H, -NH-CH_2), 3.2–3.3 (m, 2H, $H_{5'}$, $H_{5''}$), 3.7 (s, 6H, -O-CH_3), 3.9 (m, 1H, $H_{4'}$), 4.3 (m, 1H, $H_{3'}$), 5.1 (d, 1H, H_5), 6.2 (m, 1H, $H_{1'}$), 6.8–7.3 (m, 18H **Ph**), 7.6 (d, 1H, H_6), 7.6 (t, 1H, NH), s: singlet, d: doublet, t: triplet, m: multiplet, Ph: phenyl.

Compounds **1a**, **1c**, **1d** and **1e** were obtained as described for compound **1b** starting from 5′-*O*-dimethoxytrityl-3′-*O*-(t-butyldimethylsilyl)-*N*-4-methyl-2′-deoxycytidine, 5′-*O*-dimethoxytrityl-1-3′-*O*-(t-butyldimethylsilyl)-*N*-4-propyl-2′-deoxycytidine, 5′-*O*-dimethoxytrityl-3′-*O*-(t-butyldimethylsilyl)-*N*-4-allyl-2′-deoxycytidine and 5′-*O*-dimethoxytrityl-3′-*O*-(t-butyldimethylsilyl)-*N*-4-propargyl-2′-deoxycytidine respectively. Yield **1a** 78%, R_{f1a} 0.53; yield **1c** 84%, R_{f1c} 0.59; yield **1d** 90%, R_{f1d} 0.63; yield **1e** 72%, R_{f1e} 0.64 system B.

Synthesis of phosphoramidite derivatives 2

5′-O-Dimethoxytrityl-3′-O-(2-cyanoethyl-N,N-diisopropylamidophosphite)-N-4-ethyl-2′-deoxycytidine **2b**. Compound **1b** was dried by several co-evaporations and left in a desiccator overnight. 2-Cyanoethyl-*N*,*N*-diisopropylamidochlorophosphite (0.53 mmol, 0.12 ml) was added dropwise under argon atmosphere to a magnetically stirred mixture of compound **1b** (0.35 mmol, 200 mg) and diisopropylethylamine (1.43 mmol, 0.24 ml) in anhydrous dichloroethane (4 ml) at room temperature. The phosphitylation reaction was monitored by TLC analysis. After 1 h the reaction mixture was diluted with ethylacetate and washed with 10% sodium hydrogen carbonate solution, then with saturated sodium chloride solution. The organic solution was dried over sodium sulfate and concentrated under reduced pressure. The residue was purified on a silica gel column using CH_2Cl_2/AcOEt/Et_3N (60:30:10 v/v/v) as eluent. The collected fractions containing the phosphoramidite compound **2b** were pooled and concentrated under reduced pressure. Compound **2b** was then obtained as a white powder after precipitation in cold hexane (–70°C). This compound was isolated by quick filtration and then dried in a desiccator. Yield **2b** 48% (m = 130 mg), R_{f2b} system A 0.35 and 0.33, R_{f2b} system C 0.68 and 0.61.

The phosphoramidites **2a**, **2c**, **2d** and **2e** were obtained starting from modified nucleosides **1a**, **1c**, **1d** and **1e** respectively, following the procedure described for preparation of **2b**. Yield **2a** 46%, R_{f2a} system A 0.36 and 0.32, R_{f2a} system C 0.52 and 0.43; yield **2c** 47%, R_{f2c} system A 0.39 and 0.34, R_{f2c} system C 0.73 and 0.68; yield **2d** 42%, R_{f2d} system D 0.35 and 0.30, R_{f2d} system C 0.72 and 0.67; yield **2e** 49%, R_{f2e} system A 0.35 and 0.31, R_{f2e} system C 0.75 and 0.69.

Synthesis purification and characterization of modified oligonucleotides

Chain assembly was carried out on a Pharmacia Gene Assembler on solid support CPG (controlled pore glass) functionalized with a nucleoside using phosphoramidite chemistry (14). Syntheses were performed on a 1 µmol scale using 10 µmol commercial phosphoramidite or modified phosphoramidite, prepared as previously described, per cycle with a cycle time of 10 min and a coupling time of 1.5 min for the commercial phosphoramidite.

The coupling time of the modified phosphoramidite was increased to 2.5 min. The coupling yields for modified dC phosphoramidites were as good as for the natural ones. The oligonucleotides obtained were then deblocked by treatment with concentrated ammonia overnight at 60°C. After deprotection and extraction of the organic impurities, the crude reaction products were purified by ion exchange chromatography using a Pharmacia FPLC system equipped with a DEAE column (8 µM, 100 × 10 mm) from Waters. A 25 mM Tris–HCl buffer, pH 8, in acetonitrile/water (10:90 v/v) with a linear gradient of NaCl from 0 to 0.45 M over 40 min at 1 ml/min was used as eluent. The fractions were monitored by absorption at 254 nm. $R_t \approx 25$ min for all compounds.

After desalting, the purity of all oligonucleotides described was checked by reverse phase analysis using a Lichrocart system (125 × 4 mm) packed with 5 µm Lichrospher RP 18 from Merck with a linear gradient of acetonitrile from 5 to 20% for 20 min in 0.1 M aqueous triethylammonium acetate buffer, pH 7, with a flow rate of 1 ml/min. The retention times (R_t) of oligonucleotides 5′-d(CG-AYGACGA)-3′ involving one modified C at position 4 were as follows: Y = ^{4Me}C, R_t 12.1 min; Y = ^{4Et}C, R_t 12.2 min; Y = ^{4Pr}C, R_t 13 min; Y = ^{4allyl}C, R_t 12.8 min; Y = ^{4propargyl}C, R_t 12.8 min.

Full deprotection and nucleoside composition of the modified oligonucleotides were ascertained by nuclease degradation. An aliquot of oligonucleotide was digested with snake venom phosphodiesterase (Pharmacia Biotech) and alkaline phosphatase (Boehringer) in 0.1 M Tris–HCl, pH 8.2, for 19 h at room temperature. After inactivation of the enzyme at 90°C for 2 min, the digestion products were analyzed by reverse phase chromatography using a Lichrocart system (125 × 4 mm) packed with Nucleosil 100-5 C18 from Macheray Nagel equilibrated with 0.1 M aqueous triethylammonium acetate buffer, pH 7. The column was eluted at a flow rate of 1 ml/min with 0.1 M aqueous triethylammonium acetate buffer, pH 7, for 15 min and then with a linear gradient of 0–20% acetonitrile in 0.1 M aqueous triethylammonium acetate buffer, pH 7, for 40 min. Detection was performed at 260 nm. All oligonucleotides were totally degraded to nucleosides. In each case four peaks were obtained. Comparison with natural and modified nucleoside samples allowed us to identify the different peaks. Three of them whose retention times were identical for every oligonucleotide hydrolyzed correspond to dC (R_t 3.5 min), dI (R_t 9.1min) and dG (R_t 11.5 min) respectively. The presence of dI, resulting from deamination of dA, could be due to contamination of the nucleases by adenosine deaminase. Peaks corresponding to the modified dC nucleosides were eluted with the following retention times: d^{4Me}C, R_t 7.1 min; d^{4Et}C, R_t 14.9 min; d^{4Pr}C, R_t 25.9 min; d^{4allyl}C, R_t 19.8 min; d^{4propargyl}C, R_t12.6 min.

Melting experiments

Changes in absorbance with temperature of 2 µM duplexes in 10^{-2} M sodium cacodylate buffer, pH 7, containing 1 M NaCl and 2×10^{-4} M EDTA were measured at $\lambda_{260\ nm}$ in a UVIKON 941 cell changer spectrophotometer equipped with a Huber PD 415 temperature programer connected to a cryothermostat ministat circulating water bath (Huber). Samples and references were slowly heated at a rate of 0.5°C/min from 0 to 80°C. Melting temperatures (T_m) were taken as the temperature corresponding to half-dissociation of the complexes. The T_m values were determined using the first and second derivatives. The molar extinction coefficients of the sequences were determined as described in the literature (15).

RESULTS AND DISCUSSION

Design and synthesis of modified oligonucleotides

Several approaches can be used to abolish differential binding stability of the hybrids dependent on base composition, among which is modification of one of the four natural 2′-deoxynucleosides, which then forms, with the complementary nucleoside, a base pair showing a stability very close to that of other base pairs. We can also modify both 2′-deoxynucleosides in a base pair to obtain a stability similar to that of the other natural base pairs. Another approach is based on modification of two non-complementary 2′-deoxynucleosides which hybridize with the complementary nucleoside to give base pairs with similar stabilities.

These investigations allow the design of one modified 2′-deoxycytidine, d*C, which hybridizes specifically with natural dG to give a G*C base pair whose stability is very similar to that of an AT base pair. This choice was dictated by the following criteria: it is easier to find a modified GC base pair whose stability is similar to that of an AT natural base pair than to design a modified AT base pair whose stability is close to that of a GC natural base pair; preparation of oligonucleotides containing dC analogs is simpler than that of oligonucleotides built with dG analogs; modification of only one base pair rather than both simplifies the enzymatic preparation of DNA containing one or several modified nucleosides.

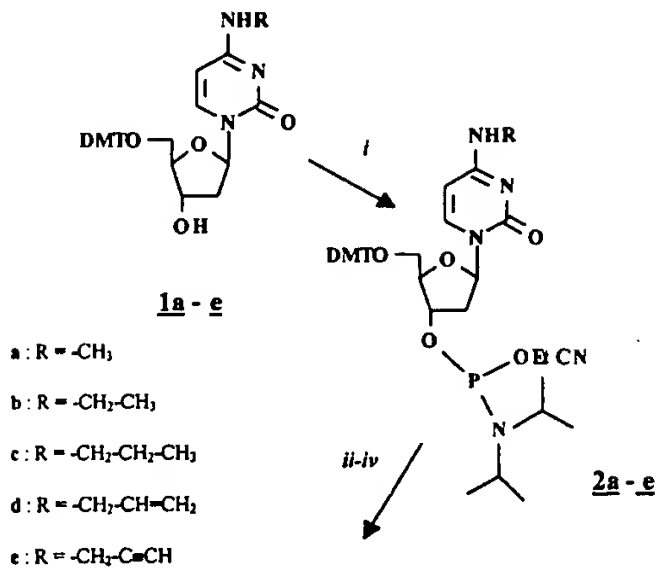
We have chosen as dC analogs the araC and dC derivatives, in which one hydrogen of the exocyclic amino group at position 4 is substituted by an alkyl group such as methyl, ethyl, *n*-propyl, allyl or propargyl groups. In fact, it is well known that replacement of one dC by ^{ara}C (16) or d^{4Me}C (17) in an oligonucleotide induces a decrease in thermal stability of the hybrid formed by this oligonucleotide and the complementary nucleic acid sequence.

Synthesis of the N-4 substituted 2′-deoxycytidine

The synthesis of 5′-*O*-dimethoxytrityl-3′-*O*-(2-cyanoethyl-*N*,*N*-diisopropylamidophosphite)-*N*-4 substituted-2′-deoxycytidine (**2a**, **2b**, **2c**, **2d** and **2e**) was carried out from commercial deoxyuridine as described in Figure 1. It consists first of protection of the 5′- and 3′-hydroxyl functions of deoxyuridine by dimethoxytrityl and t-butyldimethylsilyl groups respectively (12). Conversion of 2′-deoxyuridine to *N*-4 substituted-2′-deoxycytidine was realized by activation at the C_4 position of the protected 2′-deoxyuridine by treatment with phosphorus oxychloride in the presence of 1,2,4-triazole followed by treatment with the primary amines (13). After deprotection of the 3′-hydroxyl function by the action of tetrabutylammonium fluoride, phosphitylation at the 3′-position was achieved using 2-cyanoethyl-*N*,*N*-diisopropylamidochlorophosphite.

Hybridization properties of oligonucleotides containing a modified d*C

Preparation of duplexes in which every natural dC is replaced by a modified d*C requires a great deal of work. Therefore, studies were first carried out with 9 bp duplexes composed of a triplet repeated three times involving an AT, GC or G*C base pair at position 4 (Table 1). We chose to investigate the stabilities of duplexes composed of a tridecamer and a nonamer to mimic hybrids formed between an oligonucleotide probe and a longer nucleic acid sequence. In fact, it is well known that the presence of dangling arms at both the 3′- and 5′-positions of the



Modified oligonucleotides

DMT = dimethoxytrityl; EtCN = 2-cyanoethyl; ⟩—— = isopropyl; (*i*) 2-cyanoethyl-N,N diisopropylamidochlorophosphite, diisopropylethylamine; (*ii*) assembly of oligonucleotides; (*iii*) deprotection; (*iv*) purification.

Figure 1. Synthesis of modified oligonucleotides involving *N*-4-substituted-2′-deoxycytidine.

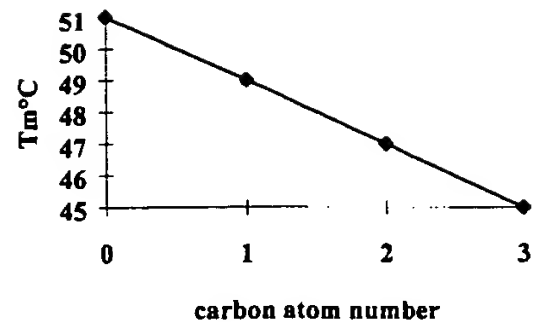


Figure 2. T_m variations of
5′-d (TTTCGTCGTCGTT) -3′
3′-d (AGCAG*CAGC) -5′
*C = $^{4-H(CH2)n}$C as a function of the carbon atom number (*n*) of the substituent.

Table 1. Melting temperatures at $\lambda_{260\ nm}$ of natural duplexes and those involving a modified *C at position 4

5′-d (T T T C G T C X T C G T T) 3′
3′-d (A G C A G Y A G C) 5′
9 8 7 6 5 4 3 2 1

Duplex	X	Y	Tm (°C)	Δ Tm/1 (°C)	Δ Tm/2 (°C)
1	A	T	46		- 5
2	G	C	51	+ 5	
3	G	4Me C	49	+ 3	- 2
4	G	4Et C	47	+ 1	- 4
5	G	4Pr C	45	- 1	- 6
6	G	4Allyl C	45.5	- 0.5	- 4.5
7	G	4Propargyl C	45.5	- 0.5	- 4.5
8	G	ara C	48	+ 2	- 3

T_m values were determined at an oligomer strand concentration of 2 μM in 10^{-2} M sodium cacodylate buffer, pH 7, containing 1 M NaCl and 2×10^{-4} M EDTA.

oligonucleotide leads to good stabilization of a duplex which varies with its length (18). The results reported in Table 1 led to the following observations.

For duplexes 3–7 involving *N*-4-substituted dC the T_m variation, measured as described in Materials and Methods, is inversely proportional to the carbon atom number of the alkyl group (Me, Et, *n*-Pr). The T_m decrease is ~2°C when one methylene group is added (Fig. 2). Duplexes 6 and 7, involving an allyl or propargyl group respectively, led to a T_m slightly higher than that of duplex 5, involving an *n*-propyl group. This difference could be due to a steric effect or the higher positive inductive effect of the *n*-propyl group versus that of the allyl or propargyl group. Note that the difference in T_m is small and within the margin of error range, which is ~±1°C. The decrease in thermal stability of duplex 3, involving one $G^{4Me}C$ base pair, compared with that of duplex 2, possessing a natural GC base pair at the same position, is small (ΔT_m ~2.5°C). Duplexes 3 and 4, involving a $G^{4Me}C$ and a $G^{4Et}C$ respectively, are thermally more stable than duplex 1, having an AT base pair at the same position (ΔT_m 3 and 1°C respectively).

Concerning double-stranded DNA involving natural nucleosides, their stabilities are highly dependent on base composition. Therefore, replacement of one AT base pair at position 4 of hybrid 1 (T_{m1} 46°C) by one GC base pair led to a 5°C increase in T_m (T_{m2} 51°C), whereas replacement of this same AT base pair by one G*C base pair (particularly *C = $d^{4allyl}C$ and $d^{4propargyl}C$) led to hybrids having a stability very similar to that of duplex 1.

Specificity of the G*C base pair

To verify the recognition specificity of dG by d*C, thermal denaturation studies were carried out with duplexes involving the mismatches XC and X*C (X = T, C or A; *C = ^{4Me}C, ^{4Et}C, ^{4Pr}C, $^{4allyl}C$, $^{4propargyl}C$ or ^{ara}C) (Table 2). The results obtained showed the following.

Specificity of G*C base pair formation was maintained. The presence of XC and X*C mismatches at position 4 led, in all cases, to a decrease of >20°C in T_m value.

The order of stabilities for duplexes involving XC and X*C mismatches was the same:
T_m (GC) >> T_m (TC) > T_m (AC) > T_m (CC);
T_m ($G^{4R}C$) >> T_m ($T^{4R}C$) > T_m ($A^{4R}C$) > T_m ($C^{4R}C$), R = Me, Et, Pr, allyl or propargyl;
T_m ($G^{ara}C$) >> T_m ($T^{ara}C$) > T_m ($A^{ara}C$) > T_m ($C^{ara}C$).

Table 2. Melting temperature values (°C) at $\lambda_{260\ nm}$ of perfect duplexes and ones containing a mismatch at position 4

5'-d (T T T C G T C X T C G T T) 3'
3'-d (A G C A G Y A G C) 5'
9 8 7 6 5 4 3 2 1

Y / X	C	^{4Me}C	^{4Et}C	^{4Pr}C	^{4Allyl}C	^{4Propargyl}C	^{ara}C
G	51	49	47	47	45.5	45.5	48
T	28.3	25	20.5	21.5	20	22.5	22
C	21	17.5	18	22	19.5	20.5	19
A	27	27	25.5	24	24	24	22

T_m values were determined on a sample concentration of 2 µM in 10^{-2} M sodium cacodylate buffer, pH 7, containing 1 M NaCl and 2×10^{-4} M EDTA.

The presence of CC or C*C mismatches has a very destabilizing effect. ^{ara}C seems to give a better discrimination than do C and ^{4R}C (Table 2).

Choice and hybridization properties of duplexes involving AT and/or G^{4Et}C base pairs

Although the previously obtained results for T_m values of duplexes **6** and **7** involving d^{4allyl}C (T_{m6} 45.5°C) and d^{4propargyl}C (T_{m7} 45.5°C) respectively were closest to that of duplex **1**, involving an AT base pair (T_{m1} 46°C), we preferred to continue our studies with d^{4Et}C, which forms a slightly more stable G^{4Et}C base pair (T_{m4} 47°C) with G than it does with the AT base pair (T_{m1} 46°C).

This modification was chosen in order to obtain modified duplexes with a not too low thermal transition and to minimize the possible steric effect of the alkyl group when duplexes were built with several contiguous d*C. We also decided to eliminate ^{4Me}C and the ^{ara}C because the first leads to a clearly more stable G^{4Me}C base pair (T_{m3} 49°C) than does the AT base pair (T_{m1} 46°C), whereas the triphosphate of the latter could not be used by polymerases (19) for the preparation of a modified DNA fragment, in contrast to N4-substitued dC derivatives, which are accepted by DNA polymerase (20). Since ^{4Me}C is present in DNA of certain thermophilic bacteria (21), we can expect that the ^{4Et}C derivative has the same physicochemical and biochemical properties. Therefore, studies were carried out with duplexes composed of 9 bp involving various base pairs ranging from nine AT base pairs to nine GC or nine G^{4Et}C base pairs (Table 3).

To determine the effect of replacing an AT base pair with a GC or G^{4Et}C base pair, studies were carried out on duplexes **9**–**11**, having the same transition of purine↔pyrimidine (duplexes **10** and **11** correspond to duplex **9** in which the AT base pairs at positions 2, 5 and 8 have been replaced with the GC and G^{4Et}C base pairs respectively), and on duplexes **2**, **12**, **13** and **14**, with

Table 3. Melting temperatures of natural and modified duplexes involving *C = ^{4Et}C

Duplex		TmfC	(AT)	(GC)
9	5'-d (T T T A T T A T T A T T T) 3' 3'-d (A T A A T A A T A) 5' 8 5 2	20	9	0
10	5'-d (T T T G T T G T T G T T T) 3' 3'-d (A C A A C A A C A) 5'	41	6	3
	3'-d (A *C A A *C A A *C A) 5' 5'-d (T T T G T T X T T G T T T) 3'			
11	X=G	27	6	3
11a	X=T	<10		
11b	X=C	<10		
11c	X=A	<10		
	3'-d (A G *C A G *C A G *C) 5' 5'-d (T T T *C G T *C X T *C G T T) 3'			
12	X=G	24	3	6
12a	X=T	<10		
13	5'-d (T T C G G C G G C G G T T) 3' 3'-d (G C C G C C G C C) 5'	59	0	9
14	5'-d (T T *C G G *C G G *C G G T T) 3' 3'-d (G *C *C G *C *C G *C *C) 5'	23	0	9
	3'-d (A G C A G C A G C) 5' 5'-d (T T T C G T C X T C G T T) 3'			
2	X=G	51.5	3	6
2a	X=T	28.5		
2b	X=C	21		
2c	X=A	27		

T_m values were determined at an oligomer strand concentration of 2 µM in 10^{-2} M sodium cacodylate buffer, pH 7, containing 1 M NaCl and 2×10^{-4} M EDTA.

different sequences. The results reported in Table 3 led us to the following observations.

Duplexes **11**, **12** and **14**, involving three, six and nine G^{4Et}C base pairs respectively, have T_m values (T_{m11} 27°C, T_{m12} 24°C, T_{m14} 23°C) very close to that of duplex **9** (T_{m9} 20°C), composed of nine AT base pairs. These results show that thermal stability of these duplexes does not depend on their base composition, contrary to what was observed with natural duplexes whose T_m values greatly increase with GC base pair number (Figure 3).

As we have previously shown, G^{4Et}C base pair formation is specific and, in fact, the T^{4Et}C mismatch (duplex **11a**), C^{4Et}C mismatch (duplex **11b**) and A^{4Et}C mismatch (duplex **11c**) led to very unstable hybrids, making it impossible to determine their T_m values (T_m <10°C).

The T_m of duplex **9** (T_{m9} 20°C), involving nine AT base pairs, is lower than that of the natural duplex having five GC base pairs, three AT base pairs and one TC mismatch (T_{m2a} 28.5°C), CC mismatch (T_{m2b} 21°C) or AC mismatch (T_{m2c} 27°C); under the same conditions the T_m of duplex **9** is higher than that of modified duplexes having five G^{4Et}C base pairs, three AT base pairs and one T^{4Et}C mismatch (T_{m11a} <10°C), C^{4Et}C mismatch (T_{m11b} <10°C) or A^{4Et}C mismatch (T_{m11c} <10°C). These results show

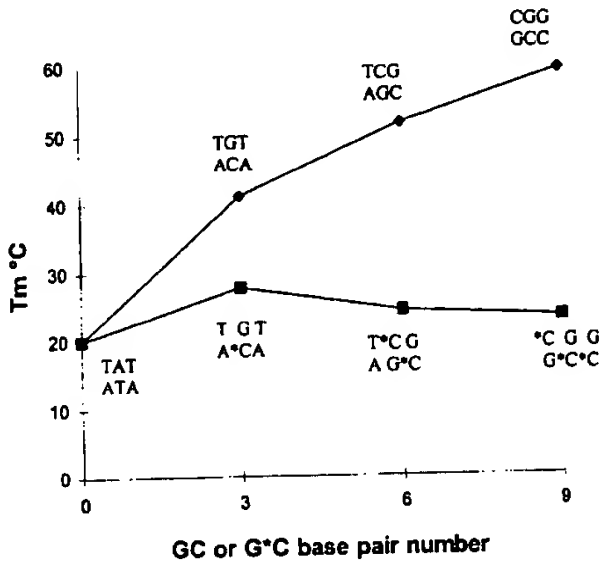


Figure 3. T_m variations of duplexes built with 9 bp

$$5'\text{-d(TT}\left|\begin{matrix}X & Y & Z\\ X' & Y' & Z'\end{matrix}\right|_3\text{TT)3}'$$

as a function of the GC (◆) or G*C (■) base pair number (*C = ^{4Et}C).

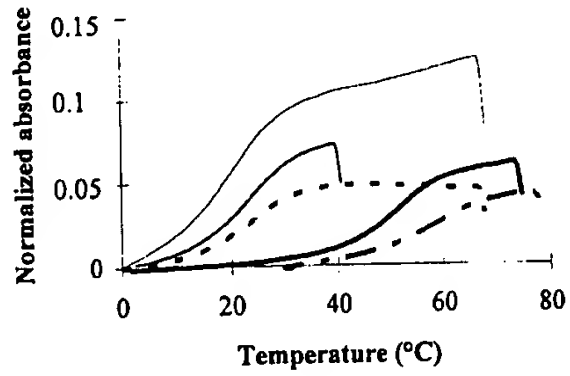


Figure 4. Melting curves of duplexes **14** (—), **12** (—), **13** (— -), **2** (___) and **9** (—) recorded at $\lambda_{260\ nm}$.

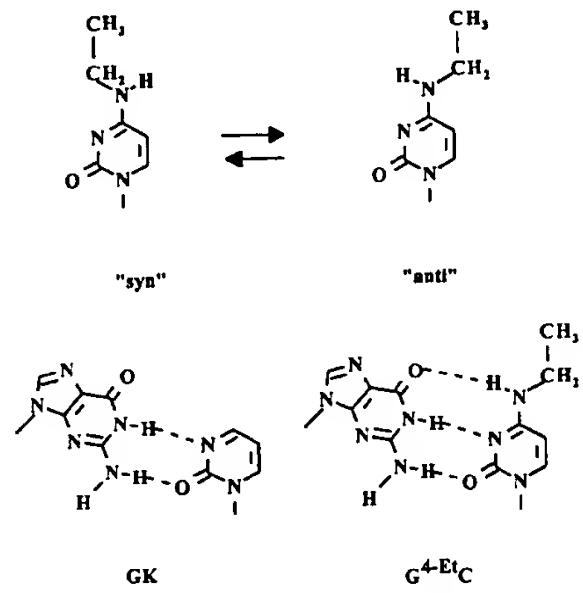


Figure 5. Structure of the '*anti*' and '*syn*' conformations of ^{4Et}C and of the GK and G^{4Et}C base pairs.

that under the hybridization conditions used it is not possible to discriminate between perfect hybrids built with AT-rich sequences and those built with GC-rich sequences involving a mismatch. This major drawback of the natural double-stranded DNA can be alleviated by using duplexes built with AT and/or G^{4Et}C base pairs.

The thermal stabilities of modified duplexes **11**, **12** and **14** are slightly higher than that of duplex **9**, having only AT base pairs. This is in accordance with previously obtained results. We note that the difference between the more stable duplex **11** (T_{m11} 27°C) and duplex **9** (T_{m9} 20°C) possessing nine AT base pairs is ~7°C. On the other hand, under the same conditions the difference in stability between natural duplexes involving nine GC (duplex **13**, T_{m13} 59°C) and nine AT base pairs (duplex **9**, T_{m9} 20°C) is ~39°C.

The T_m of duplex **14** (T_{m14} 23°C), containing nine G^{4Et}C base pairs with three ^{4Et}C^{4Et}C doublets, was as expected, similar to that of duplexes containing nine AT base pairs (T_{m9} 20°C), six AT and three G^{4Et}C (T_{m11} 27°C) or three AT and six G^{4Et}C (T_{m12} 24°C) respectively, where the d^{4Et}C residues were not consecutive (Table 3 and Fig. 3). Consequently, substitution at position 4 of the C by an ethyl group does not lead to notable steric disturbance. These results were confirmed by the high cooperativity of dissociation of modified duplexes **12** and **14** into single strands, which was similar to that of natural duplexes **2** and **13** (Fig. 4).

For duplexes **9**, **12** and **14**, involving nine AT base pairs, three AT base pairs and six G^{4Et}C base pairs and nine G^{4Et}C base pairs respectively, the increase in absorbance recorded at $\lambda_{260\ nm}$ at the time of dissociation into single strands decreased when G^{4Et}C base pair number increased. The same phenomenon was also observed with natural duplexes **2** (three AT base pairs and six GC base pairs) and **13** (nine GC base pairs). These results suggest that the GC and G^{4Et}C base pairs have similar physicochemical properties. It is highly likely that ^{4Et}C forms a G^{4Et}C base pair with G having three hydrogen bonds, since the G^{4Et}C base pair stability would otherwise be weaker, as was shown for a GK base pair (K being 2-pyrimidinone), which has two hydrogen bonds (22) and is isomorphous with the GC base pair (Fig. 5).

In single-stranded structures the ethyl group in ^{4Et}C has '*syn*' and '*anti*' conformations, the balance of which can be quickly shifted to the '*anti*' configuration by hybridization with G. In fact, variations in optical density as a function of temperature recorded during duplex dissociation into single strands or during association of the latter into duplexes are identical (results not shown).

Several hypotheses could be envisaged to explain the G^{4Et}C base pair stability decrease compared with that of the natural GC base pair.

Substitution of one hydrogen of the amino group at position 4 of C with one alkyl group (R = Me, Et, Pr, ...) induces a positive inductive effect higher than that of the hydrogen, which makes the other hydrogen less electrophilic leading to a weakening of the hydrogen bond between the H atom at position 4 of C and the oxygen atom at position 6 of G.

The presence of a lipophilic alkyl group whose size is relatively large could modify the conformation and/or hydration of duplexes.

All the results obtained show that the stability of modified duplexes built with AT and/or G^{4Et}C base pairs is not dependent on their AT/G^{4Et}C content ratio. This new system may be very

useful for DNA sequencing by hybridization, allowing discrimination between perfect hybrids and those involving mismatches.

CONCLUSION

We have proven, for the first time, that the stability of duplexes made with oligonucleotides of a given length built with AT and/or $G^{4Et}C$ base pairs is independent of their base content in a classical buffer solution. Specificity of the $G^{4Et}C$ base pair is maintained and the cooperativity of modified duplex dissociation is similar to that of natural DNA duplexes involving natural base pairs. These very useful properties make possible the employment of this new system in reverse hybridization approaches, using a large number of sequences immobilized as a two-dimensional matrix for simple and fast analysis of nucleic acid sequences, or in biochemical techniques, such as random priming, for the preparation of DNA and labeling DNA fragments by an enzymatic method. Work is currently in progress to confirm the hybridization properties of such modified duplexes with large oligonucleotide sequences, to determine the influence of the sequence on thermal stability and to prepare a DNA target fragment using $d^{4Et}C$ triphosphate and DNA polymerases.

ACKNOWLEDGMENT

This work was supported by a BioMed contract (gene-CT 93-0009).

REFERENCES

1 Southern,E.M. (1989) International patent WO 89/10977.
2 Drmanac,R., Labat,I., Bruckner,I. and Crkvenjakov,R. (1989) *Genomics*, **4**, 114–128.
3 Krapko,K.R., Lysov,Y.P., Khorlin,A.A., Ivanov,I.B., Yershov,A.D., Vasilenko,S.K., Florentiev,V.L. and Mirzabekov,A.D. (1991) *J. DNA Sequencing Mapping*, **1**, 375–388.
4 Melchoir,W.B.,Jr and Von Hippel,P.H. (1973) *Proc. Natl. Acad. Sci. USA*, **70**, 298–302.
5 Wood,W.I., Gitschier,J., Ladky,L.A. and Lawn,R.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1585–1588.
6 Jacobs,K.A., Rudersdorf,R., Neill,S.D., Dougherty,J.P., Brown,E.L. and Frisch,E.F. (1988) *Nucleic Acids Res.*, **16**, 4637–4650.
7 Maskos,U. and Southern,E.M. (1993) *Nucleic Acids Res.*, **21**, 4663–4669.
8 Ricelli,P.V. and Benight,A.S. (1993) *Nucleic Acids Res.*, **21**, 3785–3788.
9 Innis,M.A., Gelfand,D.H., Sninsky,J.J. and White,T.J. (Eds) (1990) *PCR Protocols: A Guide to Methods and Applications*. Academic Press, New York, NY.
10 Hoheisel,J.D. and Lehrach,H. (1990) *FEBS Lett.*, **274**, 103–106.
11 Hoheisel,J.D. (1996) *Nucleic Acids Res.*, **24**, 430–432.
12 Fritz,H.J., Frommer,W.B., Kramer,W. and Werr,W. (1982) In Gassen,H.G. and Lang,A. (Eds), *Chemical and Enzymatic Synthesis of Gene Fragmants. A Laboratory Manual*. Verlag Chemie, Weinheim, Germany, pp. 43–52.
13 Sung,W.L. (1981) *Nucleic Acids Res.*, **9**, 6139–6151.
14 Beaucage,S.L. and Caruthers,M.H., (1981) *Tetrahedron Lett.*, **22**, 1859–1862.
15 Fasman,G.D. (Ed.) (1975) *Handbook of Biochemistry and Molecular Biology, Nucleic Acids*, 3rd Edn. Vol. 1. CRC Press, Cleveland, Ohio, pp. 589.
16 Giannaris,P.A. and Damha,M.J. (1994) *Can. J. Chem.*, **72**, 909–918.
17 Butkus,V., Klimasauskas,S., Petrauskiene,L., Maneliene,Z., Janulaitis,A., Minchenkova,L.E. and Schyolkina,A.K. (1987) *Nucleic Acids Res.*, **15**, 8467–8478.
18 Williams,J.C., Case-Green,S.C., Mir,K.U. and Southern,E.M., (1994) *Nucleic Acids Res.*, **22**, 1365–1367.
19 Mikita,T. and Beardsley,G.P. (1994) *Biochemistry*, **33**, 9195–9208.
20 Li,S., Haces,A., Stupar,L., Gebeyehu,G. and Pless,R.C. (1993) *Nucleic Acids Res.*, **21**, 2709–2714.
21 Ehrlich,M., Norris,K.F., Wang,R.Y.-H., Kuo,K.C. and Gehrke,C.W. (1986) *Biosci. Rep.*, **6**, 387–393.
22 Gildea,B. and McLaughlin,L.W. (1989) *Nucleic Acids Res.*, **17**, 2261–2281.

J. Mol. Biol. (1996) 255, 589–603

JMB

An Approach to Random Mutagenesis of DNA Using Mixtures of Triphosphate Derivatives of Nucleoside Analogues

Manuela Zaccolo[1]*, David M. Williams[2], Daniel M. Brown[2] and Ermanno Gherardi[1]

[1]*ICRF Cell Interactions Laboratory, MRC Centre Hills Road, Cambridge CB2 2QH, UK*

[2]*MRC Laboratory of Molecular Biology MRC Centre, Hills Road Cambridge, CB2 2QH, UK*

**Corresponding author*

We describe a new method for random mutagenesis of DNA based on the use of a mixture of triphosphates of nucleoside analogues. The method relies on DNA amplification *in vitro* with *Taq* polymerase and in the presence of the 5'-triphosphates of 6-(2-deoxy-β-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one(dP) and of 8-oxo-2'deoxyguanosine (8-oxodG). The newly synthesised triphosphate derivative of dP (dPTP) is an excellent substrate for *Taq* polymerase (K_m = 22 μM *versus* K_m = 9.5 μM for TTP); it is incorporated in place of TTP and, with a ~fourfold lower efficiency, in place of dCTP. After 30 cycles of DNA amplification, equimolar mixtures of the four normal dNTPs and dPTP yield the following frequencies of the four transition mutations: A→G (4.4×10^{-2}), T→C (4.3×10^{-2}), G→A (1.1×10^{-2}) and C→T (1.0×10^{-2}). The triphosphate derivative of 8-oxodG (8-oxodGTP) is incorporated opposite template adenine and yields two transition mutations (A→C and T→G) at frequencies of 0.8×10^{-2} and 1.2×10^{-2} respectively. Reaction mixtures containing dPTP and 8-oxodGTP results in both dP and 8-oxodG-induced mutations and an extensive array of codon changes in the absence of insertions and deletions. The method described differs from previous mutagenesis procedures in three respects: (1) it enables very high frequencies of base substitutions (up to 1.9×10^{-1}), (2) it allows control of the mutational load *via* the number of DNA amplification cycles and (3) it yields both transition and transversion mutations. The procedure may find application in the generation of libraries of DNA and protein mutants from which species with improved or novel activities may be selected.

Keywords: pyrimidine analogue; dP; 8-oxodeoxyguanosine; PCR; random mutagenesis; protein engineering

Introduction

Protein engineering emerged in the early eighties with the advent of site-specific mutagenesis, i.e. with the ability to introduce specific amino acid substitutions in protein sequences through mutagenic oligonucleotides (Hutchinson *et al.*, 1978; Gillam & Smith, 1979a,b). Over the last 15 years, this approach has yielded a wealth of data on the role of individual residues on the structure and function of a variety of proteins and has enabled the production of enzymes (Winter *et al.*, 1982; Estell *et al.*, 1985), antibodies (Neuberger *et al.*, 1984) and transporters (Looker *et al.*, 1992) with improved or novel properties. An extended application of this approach has allowed the transfer of specific functions from one protein to another, the recipient protein being, typically, a homologue of the donor protein. This has been exemplified successfully with grafts of antigen-binding loops in antibodies (Jones *et al.*, 1986) and of allosteric functions in transporter proteins such as haemoglobin (Komiyama *et al.*, 1995). In the latter example, 12 different residues located in two different subunits (α_1 and β_2) had to be substituted in order to transplant a bicarbonate binding site into human haemoglobin. Clearly, an obligatory requirement for success in these experiments is high-resolution molecular structures of the donor and/or recipient proteins or of close homologues.

Abbreviations used: 8-oxodG, 8-oxo-2'-deoxyguanosine; dP, 6-(2-deoxy-β-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-c][1,2]oxazin-7-one.

0022-2836/96/040589-15 $12.00/0

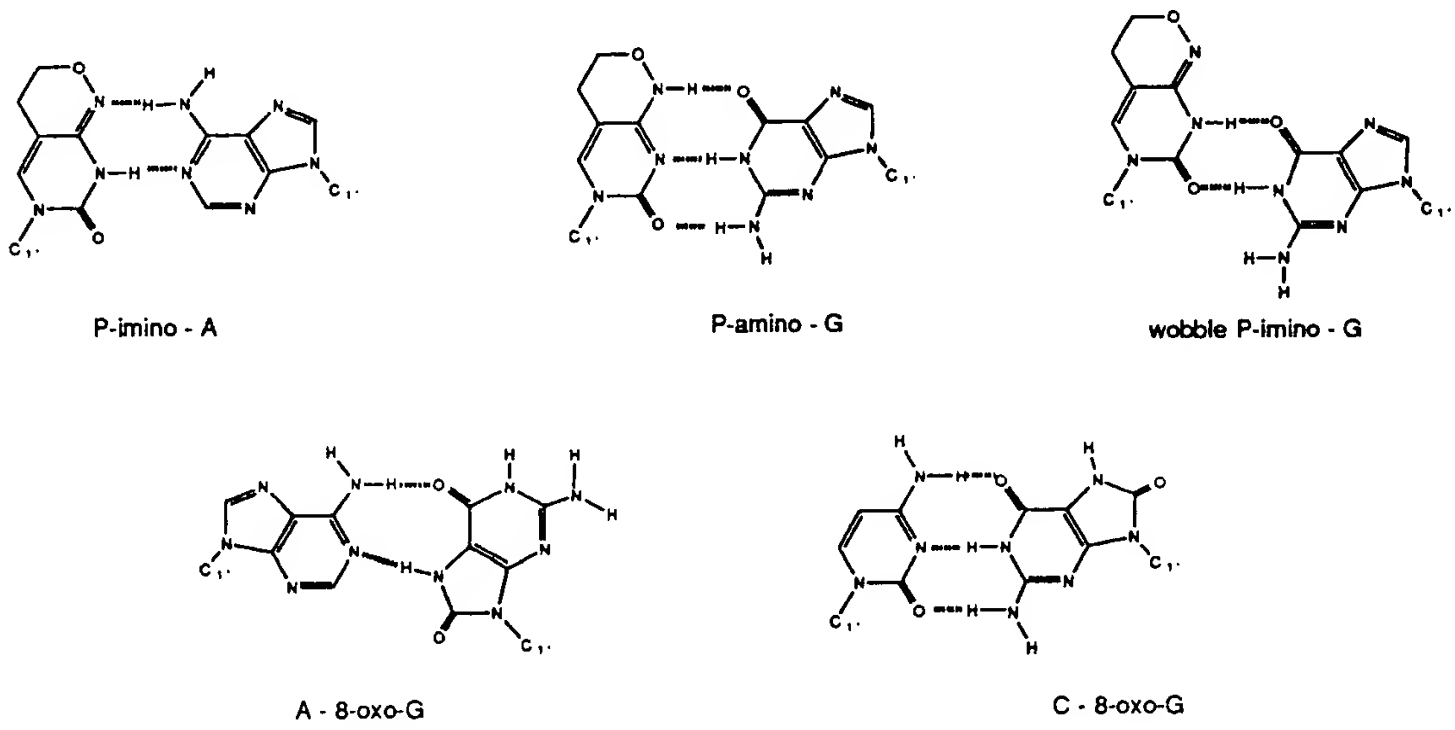


Figure 1. Base-pairing of P with A and G and 8-oxoG with A and C.

A different and complementary approach to protein engineering involves the generation of random mutants followed by selection of the ones with improved or new phenotypes, an approach that mimics an important aspect of natural protein evolution. While it is unlikely that this approach can generate functions that require the concurrent replacement of a large number of amino acids (such as the bicarbonate effect in haemoglobin), it has nevertheless considerable potential for the development of proteins with improved properties.

The power of this experimental approach is well illustrated by comparison with natural somatic hypermutation of antibody genes. Following antigen binding, antibody-producing cells undergo proliferation as well as targeted hypermutation of the variable (V) segments. The latter process, coupled with stringent selection for high-affinity variants, generates in a short period of time antibodies with binding affinities 100 to 1000-fold higher than those encoded by their germ line counterparts (reviewed by Berek & Milstein, 1987).

The successful adaptation of such a strategy to protein engineering *in vitro* is conceptually attractive but it depends critically on the availability of efficient random mutagenesis procedures and the ability to clone large libraries of mutants and select the species with the desired phenotype. In the last few years considerable progress has been made towards the construction of large libraries, at least in prokaryotes (Dower *et al.*, 1988; Griffiths *et al.*, 1994) and the development of suitable selection systems, notably through the advent of the phage display technology (Smith, 1985). There is no question, however, that there remains a requirement for controlled and efficient random mutagenesis procedures.

We took the view that for random mutagenesis of extended DNA sequences a PCR-based method would be advantageous, utilising the triphosphates of mutagenic nucleosides as substrates. The requirements would be that the substrates should be stable under the conditions of PCR cycling, well incorporated, be of high mutagenic efficiency and between them direct both transition and transversion point mutagenic changes. There are many base analogue transition mutagens whose 5'-triphosphates have been described. These include N^4-hydroxy-2'-deoxycytidine (Müller *et al.*, 1978), 2-aminopurine-2'-deoxyriboside (Grossbergher & Clough, 1981), 5-bromo-2'-deoxyuridine (Mott *et al.*, 1984), O^6-methyl-2'-deoxyguanosine (Eadie *et al.*, 1984; Snow *et al.*, 1984), N^4-amino-2'-deoxycytidine (Negishi *et al.*, 1985), N^4-methoxy-2'-deoxycytidine (Singer *et al.*, 1984; Reeves & Beattie, 1985), N^6-hydroxy-2'-deoxyadenosine (Abdul-Masih & Bessman, 1986) and 5-hydroxy-2'-deoxycytidine and -uridine (Purmal *et al.*, 1994). However, none of these have been shown to fulfil all the above criteria. In order to be effective in inducing transition mutations, the nucleoside analogue has a requirement for a tautomeric constant (ratio of imino/amino forms) close to unity. N^4-Amino-2'-deoxycytidine has a tautomeric constant between 0.1 and 10 (Brown *et al.*, 1968) and its triphosphate is well incorporated as an analogue of dCTP and TTP by *Escherichia coli* DNA polymerase I (Negishi *et al.*, 1985). It is however relatively unstable. N^4-Hydroxy-2'-deoxycytidine and N^4-methoxy-2'-deoxycytidine both display tautomeric constants between 10 and 30 (Brown *et al.*, 1968; Morozov *et al.*, 1982) and thus are able to form Watson-Crick base-pairs with both A and G. However, in the imino forms of these analogues, the amino substituent prefers to adopt the *syn* conformation (with respect to N3) (Morozov *et al.*, 1982; Shugar

et al., 1976), which shields the hydrogen bonding face of the molecule and thereby reduces duplex stabilities (Anand *et al.*, 1987). For this reason, although they are incorporated into DNA by various DNA polymerases, they are poor substrates (Singer *et al.*, 1984; Reeves & Beattie, 1985). The nucleoside dP (Figure 1) is an analogue of N^4-methoxy-2'-deoxycytidine, which, by virtue of its bicyclic ring structure, is restricted to the *anti* conformation. It thus forms stable base-pairs with both A and G (Figure 1) comparable to T and C as evidenced by the melting behaviour of oligonucleotides containing them (Kong Thoo Lin & Brown, 1989). NMR data (Nedderman *et al.*, 1993) and crystallography (Moore *et al.*, 1995) show that in duplexes P forms a Watson-Crick pair with both A and G; with the latter this is in rapid chemical exchange with a wobble pairing in solution (Stone *et al.*, 1991; Nedderman *et al.*, 1993). As it is widely accepted that base-pair geometry is important in substrate-template recognition by DNA polymerases, we expected that the triphosphate of the P deoxynucleoside (dPTP) would be a good substrate for *Taq* polymerase.

In contrast, there are very few reports of nucleoside triphosphate analogues capable of transversion mutations (Purmal *et al.*, 1994; Pavlov *et al.*, 1994; Singer *et al.*, 1986, 1989). Of these only the triphosphate of 8-oxo-2'-deoxyguanosine (8-oxodGTP) appears to be a substrate for DNA polymerases adequate for mutagenesis (Purmal *et al.*, 1994; Pavlov *et al.*, 1994). It base-pairs with both adenine and cytosine (Figure 1); the analogue assumes the normal *anti* conformation when paired with cytosine (Oda *et al.*, 1991), but adopts the *syn* conformation when forming a Hoogsteen pair with adenine (Kouchakdjian *et al.*, 1991) and therefore elicits A → C transversions (Shibutani *et al.*, 1991; Maki & Sekiguchi, 1992; Cheng *et al.*, 1992).

Here we report on an approach to random mutagenesis of DNA using dPTP and 8-oxodGTP. The two analogues are efficiently incorporated into DNA *in vitro* by the thermostable *Taq* polymerase and this allows for point mutations to be introduced with high efficiency into the replicated target sequence by PCR. We also show that their frequency can be determined by the number of amplification cycles.

Results

Synthesis of dPTP and 8-oxodGTP

The triphosphate of 8-oxodGTP was prepared from dGTP using the literature procedure (Mo *et al.*, 1992). The triphosphate analogue dPTP was prepared from the free nucleoside (Kong Thoo Lin & Brown, 1989) in 34% yield by phosphorylation with phosphoryl chloride, followed by reaction with pyrophosphate as described (Ludwig, 1981). Both triphosphates were purified by reverse phase HPLC; dPTP was initially purified by anion exchange chromatography.

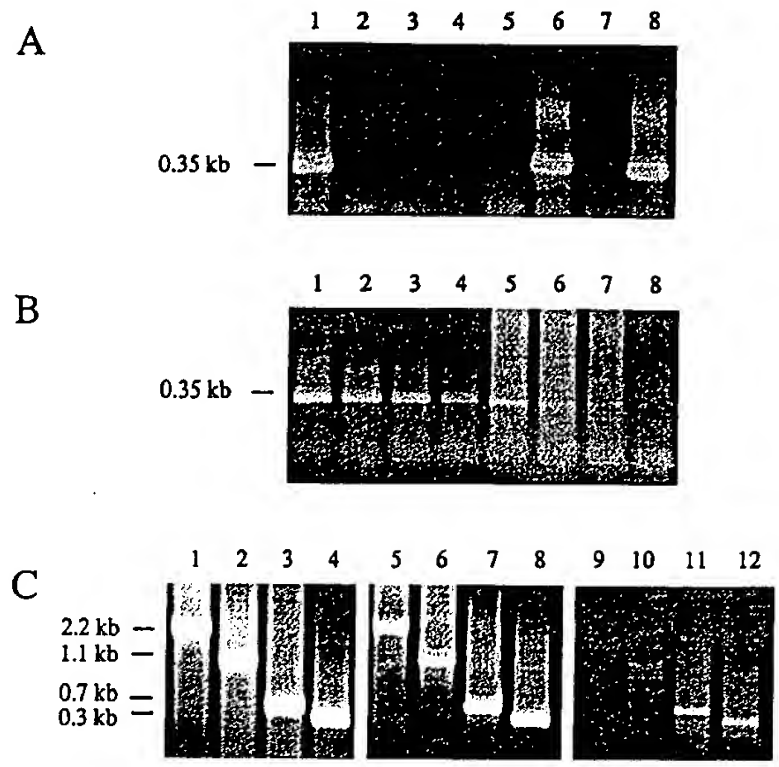


Figure 2. Incorporation into DNA and extension of dPTP (A) and 8-oxodGTP (B) by *Taq* polymerase. As template DNA the V region of the human antibody MH22 was used. A, The PCR reaction mixtures contained the following nucleoside triphosphates: dATP, dGTP, dCTP, TTP (1); dGTP, dCTP, TTP, dPTP (2); dATP, dCTP, TTP, dPTP (3); dCTP, TTP, dPTP (4); dATP, dGTP, TTP, dPTP (5); dATP, dGTP, dCTP, dPTP (6); dATP, dGTP, dPTP (7); dATP, dGTP, dCTP, TTP, dPTP (8). All dNTPs were at 500 μM except in samples 4 and 7 in which dPTP was at 1 mM. B, The PCR reaction mixture included: dATP, dCTP and TTP at 500 μM. Samples 1 to 4 contained dGTP at 50 μM, 25 μM 12.6 μM and 6.25 μM, respectively and 8-oxodGTP at 500 μM. Samples 5 to 8 contained the same concentrations of dGTP as samples 1 to 4 but no 8-oxodGTP. C, Amplification by PCR of different target genes in the presence of the four natural dNTPs (lanes 1 to 4); equimolar concentration of the four normal dNTPs and dPTP (lanes 5 to 8); equimolar concentrations of the four normal dNTPs, dPTP and 8-oxodGTP (lanes 9 to 12). The template DNA was: human macrophage stimulating protein (MSP; lanes 1, 5 and 9); human connexin 43 (lanes 2, 6 and 10); human connexin 31 (lanes 3, 7 and 11); the ζ chain of human CD3 (lanes 4, 8 and 12). The size of the fragments is indicated on the side.

Incorporation and extension of dPTP and 8-oxodGTP in template DNA by *Taq* polymerase

For the purpose of DNA mutagenesis we used *Taq* polymerase because of the advantage that a thermostable enzyme has when performing multiple cycles of DNA synthesis and because of its lack of proof-reading activity (Tindall & Kunkel, 1988).

The ability of *Taq* polymerase to use dPTP as a substrate in DNA synthesis was assayed by amplifying a target gene in a PCR reaction mixture in which one or two of the normal dNTPs were replaced by the analogue. The results are shown in Figure 2A. In experiments in which a 350 bp DNA fragment was amplified, dPTP was able to replace TTP (lane 6) yielding amounts of PCR product

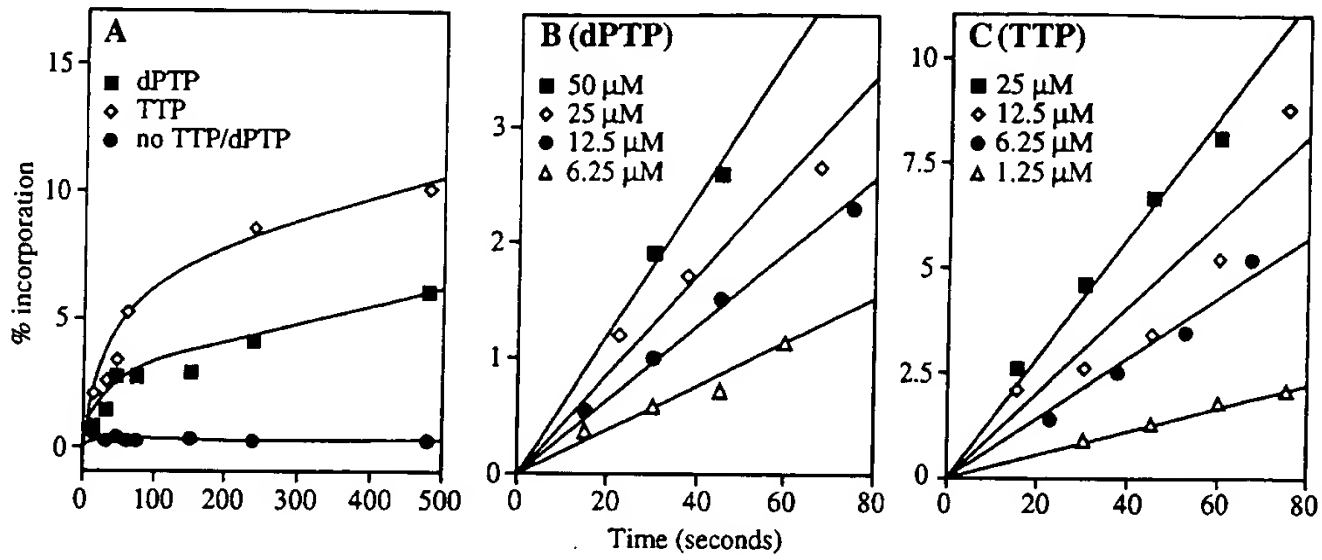


Figure 3. A, Time course of DNA synthesis in the presence of 12.5 µM [^{32}P]dCTP, dATP, dGTP and TTP (open diamonds) or dPTP (filled squares). Primed M13mp18 was used as a template for DNA synthesis in the presence of 0.3 unit *Taq* polymerase. B and C show the rate of DNA synthesis during the first 80 seconds of the reaction in the presence of 12.5 µM dATP, dGTP and [^{32}P]dCTP and the indicated concentrations of dPTP (B) and TTP (C).

comparable to those obtained with the four normal dNTPs (lane 1). A much lower yield of DNA was obtained when dPTP was used in place of dCTP (lane 5). The analogue was not able to replace dATP (lane 2), dGTP (lane 3), both dATP and dGTP (lane 4), or both dCTP and TTP (lane 7). When we substituted a normal dNTP with dPTP the yield of full length product after PCR was influenced by the length and sequence of the DNA target (data not shown). This was never the case when dPTP was added at an equimolar concentration to the four normal dNTPs (see below).

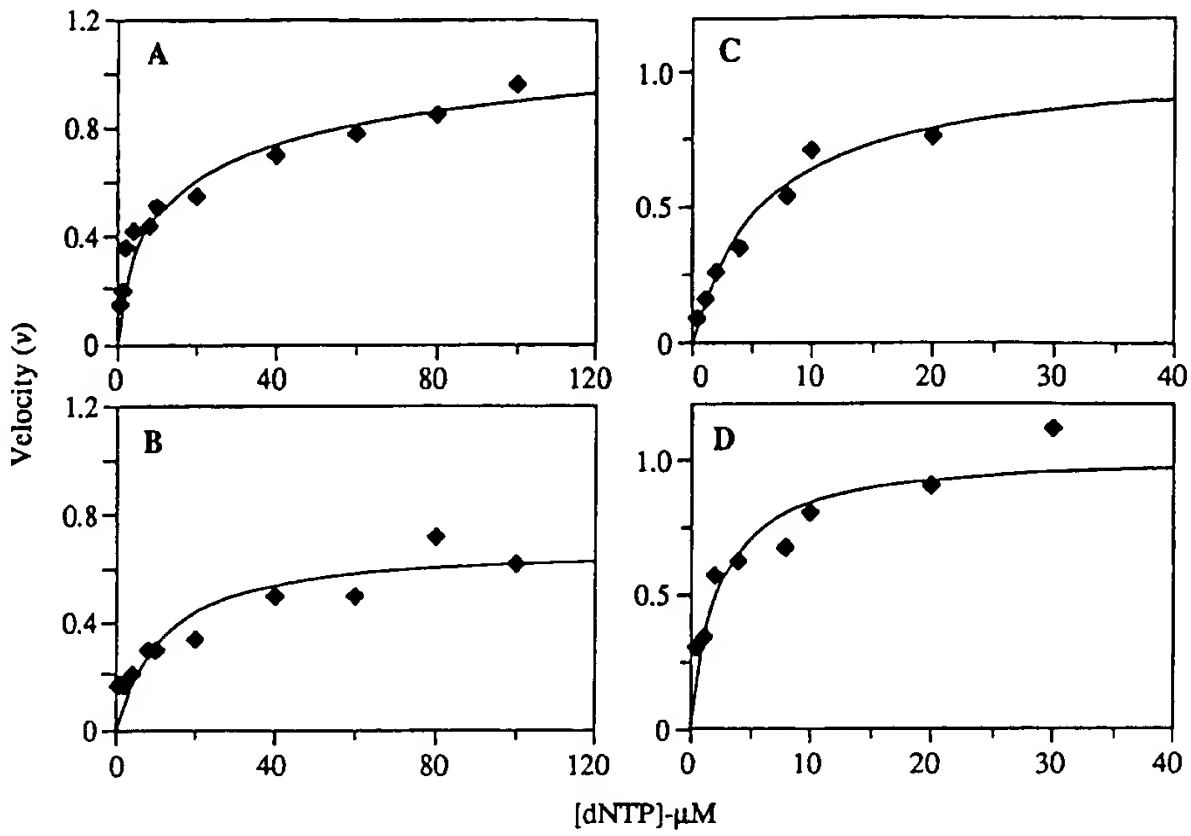
In contrast, none of the normal dNTPs could be replaced by 8-oxodGTP in a PCR reaction even with short template sequences (350 bp; data not shown). However, incorporation of 8-oxodGTP and chain extension was demonstrated by using 500 µM dATP, TTP and dCTP and limiting concentrations of dGTP in the presence of high concentrations of the base analogue (Figure 2B). Lanes 5 to 8 show the PCR products obtained in the presence of decreasing concentrations of dGTP (50 µM, 25 µM 12.6 µM and 6.25 µM respectively). The addition of 500 µM 8-oxodGTP to the same PCR reaction mixtures gave an increased yield of DNA (compare lanes 2 to 4 with lanes 6 to 8).

In order to assess the general applicability of the method, we amplified by PCR DNA fragments of different sizes in the presence of dPTP or a combination of dPTP and 8-oxodGTP (Figure 2C). Using equimolar concentrations of dPTP and the four normal dNTPs efficient amplification of target DNA sequences from 0.3 to 2.2 kb in size could be achieved (lanes 5 to 8), which was comparable to that obtained in the absence of base analogues (lanes 1 to 4). When equimolar concentrations of dPTP, 8-oxodGTP and the four normal dNTPs were used, the yield of DNA decreased, especially in the case of longer target DNA fragments (lanes 9 to 12).

Kinetics of incorporation of dPTP by *Taq* polymerase

In order to evaluate the performance of dPTP as a substrate for *Taq* polymerase, its rate of incorporation was analysed and compared with TTP since initial experiments indicated that its properties best resembled those of this natural triphosphate. Figure 3A shows the rate of DNA synthesis in the presence of dATP, dCTP and dGTP plus TTP or dPTP. DNA synthesis was measured by the incorporation of [α-^{32}P]dCTP using a primed M13 template at 72°C. Incorporation increased linearly in the first 80 seconds when either dPTP or TTP was present. In order to calculate rates of incorporation for different concentrations of substrate, time points were chosen over intervals in which both triphosphate derivatives gave a linear rate of synthesis (Figure 3B and C). Concentrations lower than 50 µM had to be used for dPTP because with higher concentrations the rate of DNA synthesis did not increase linearly with time. For TTP, concentrations between 1.25 and 25 µM were used to obtain measurable differences in rates of incorporation over time (Figure 3C). The apparent K_m values for TTP and dPTP were determined by analysing the experimental data by the direct linear plot method (Eisenthal & Cornish-Bowden, 1974). The apparent K_m for dPTP under these experimental conditions was 22 µM, whilst that for TTP was 9.25 µM. The value for dPTP thus compares favourably with those reported in the literature (Kong *et al.*, 1993) for the four natural dNTPS (14 to 17 µM).

In order to compare the relative efficiencies of insertion of dPTP opposite template adenine and guanine residues, respectively, we adopted the procedure of Boosalis *et al.* (1987) for the determination of steady state kinetics using one of two primed synthetic oligonucleotide templates



^{1} ^{23}
5'-d(GGCCTT GATATTCACAAACGAAT) 3' PRIMER
3'-d(CCGGAACTATAAGTGTTT GC TT ACCATTCT) 5' TEMPLATE 1 (dPTP/dTTP)
1 2 3
3'-d(CCGGAACTATAAGTGTTT GC TT ACCGTTCT) 5' TEMPLATE 2 (dPTP/dCTP)

Figure 4. Top: Plots of initial velocities *versus* [dNTP] for the incorporation by *Taq* polymerase of dPTP opposite A (A), TTP opposite A (B), dPTP opposite G (C), dCTP opposite G (D). Below: Primer and templates used in experiments. *v* is as defined by Boosalis *et al.* (1987).

(Figure 4). The ^{32}P-labelled primer in each case was extended by the incorporation of dGTP at two positions, followed by dPTP (template 1 and 2), TTP (template 1) or dCTP (template 2). Separation of the products by PAGE followed by quantitation of the radioactivity using a PhosphorImager allowed the determination of the initial velocities (Boosalis *et al.*, 1987). Due to the very high extension rate of *Taq* polymerase, the kinetic parameters were determined at 55°C. The velocities for the insertion of the particular triphosphate opposite template V_{max} and K_m values (μM) for particular insertions were determined from non-linear regression fitting to the Michaelis-Menton equation.

Plots of *v versus* substrate concentration [S] are illustrated for the four possibilities PA, PG, TA and CG in Figure 4A to D and the kinetic parameters and catalytic efficiencies (V_{max}/K_m) are given in Table 1. The results indicate that dPTP is virtually indistinguishable from TTP in terms of its recognition by *Taq* polymerase. Furthermore, it is incorporated approximately three times more efficiently opposite template adenine rather than guanine residues.

K_m values have been reported for 8-oxodGTP with *E. coli* DNA polymerase I Klenow fragment (Purmal *et al.*, 1994) using a procedure analogous to that described here. Values of 63 and 58 μM for insertion opposite C and A, respectively, were obtained at 37°C and compare with an average value of approximately 1 μM for the normal dNTPs (Purmal *et al.*, 1994). In addition, the analogue is a substrate for the thermostable *Tth* DNA polymerase and has been shown to generate A → C transversions at a rate of about 1% (Pavlov *et al.*, 1994).

Table 1. Kinetic parameters for dPTP with *Taq* polymerase

Template	Substrate	V_{max} (rel)	K_m (μM)	V_{max}/K_m (M^{-1})	Relative efficiency
A	dPTP	0.86 ± 0.06	5.2 ± 1.5	16.5×10^4	0.99
G	dPTP	0.69 ± 0.08	12.1 ± 4.8	5.7×10^4	0.11
A	TTP	1.02 ± 0.11	6.1 ± 1.5	16.7×10^4	1.00
G	dCTP	1.01 ± 0.09	2.03 ± 0.68	49.8×10^4	1.00

The data are derived from Figure 4. For conditions refer to Materials and Methods. V_{max} (rel) is as described by Boosalis *et al.* (1987).

PCR mutagenesis with dPTP, 8-oxodGTP and their mixtures

PCR mutagenesis was performed by amplifying a target DNA fragment with *Taq* polymerase in the presence of equimolar concentrations of the four normal dNTPs, dPTP and/or 8-oxodGTP. The product of this first PCR was subjected to a second PCR in the presence of the four natural dNTPs in order to eliminate the base analogues from the target DNA before cloning and transforming the DNA into *E. coli*. In this way the pattern of mutation was not influenced by the DNA repair mechanisms of the host cell.

Figure 5 shows the results of a series of mutagenesis experiments in which the following equimolar nucleotide mixtures were used: the four normal dNTPs together with either dPTP (Figure 5A), or 8-oxodGTP (Figure 5B) or dPTP and 8-oxodGTP (Figure 5C). Similar experiments were carried out on a second target gene with comparable results (data not shown). DNA amplification reactions were carried out for a variable number of cycles, as indicated in Figure 5 by the number preceding the point in the clone designation. The data show that a significant number of point mutations were generated in the target gene under the three experimental conditions tested although dPTP clearly proved to be a much more efficient mutagen than 8-oxodGTP, as expected. The data also clearly show that the number of mutations increased as a function of the number of cycles used for the DNA amplification reaction. When the frequency of mutations was plotted against the number of PCR cycles (Figure 6) a linear relation was apparent both in the case of 8-oxodGTP and for the mixture of dPTP and 8-oxodGTP at least up to 30 cycles. In the case of dPTP the relation was linear for the first 20 cycles. For low numbers of cycles, the combination of the two triphosphate analogues appeared to produce a total number of mutations lower than that produced by dPTP alone, although the DNA produced in such reactions contained both dP-induced and 8-oxodG-induced mutations (see below). Although the clones sequenced after different numbers of PCR cycles were obtained from separate PCR reactions and mutations appeared to accumulate over the entire gene sequence, it is interesting to note that bases at particular positions were mutated more frequently than others (Figure 5).

The total number of bases sequenced in the cloned inserts and the mutations generated by dPTP, 8-oxodGTP and their combination are listed in Table 2. The pattern of mutations produced by dPTP, 8-oxodGTP and their combination is shown in Figure 7. Thus, of the mutations generated by dPTP, 46.6% were A → G, 35.5% were T → C, 9.2% were G → A and 8% were C → T. The major mutational events (A → G and T → C transitions) resulted from the preferential incorporation of dPTP opposite A in either strand and subsequent pairing of the incorporated dP with G. Incorporation of dPTP opposite G in either strand and subsequent pairing of dP with A (G → A and C → T transitions) occurred albeit at lower frequency (~1/4) as expected from the results shown in Figure 2 and Table 1. In addition to the four transitions mentioned above, one T → G and two A → T transversions were found out of 4093 bp sequenced.

In the mutants generated with 8-oxodGTP two types of transversion mutations were present: A → C (38.8%) and T → G (59%) (Figure 7). These derive from the misincorporation of 8-oxodGTP opposite A in either template strand (Shibutani *et al.*, 1991). One C → A transversion was found out of 5463 bp sequenced. This mutation might be due to incorporation of 8-oxodGTP opposite C in the template followed by misincorporation of dATP opposite template 8-oxodG during subsequent replication. This mutagenic mechanism for 8-oxodGTP has been previously reported to occur when 8-oxodGTP completely substitutes for dGTP. However, the frequency of C → A transversions was 40 times lower than that of the A → C transversions (Cheng *et al.*, 1992). A very small number of additional mutations were also found: two A → G transitions and one G → A transition.

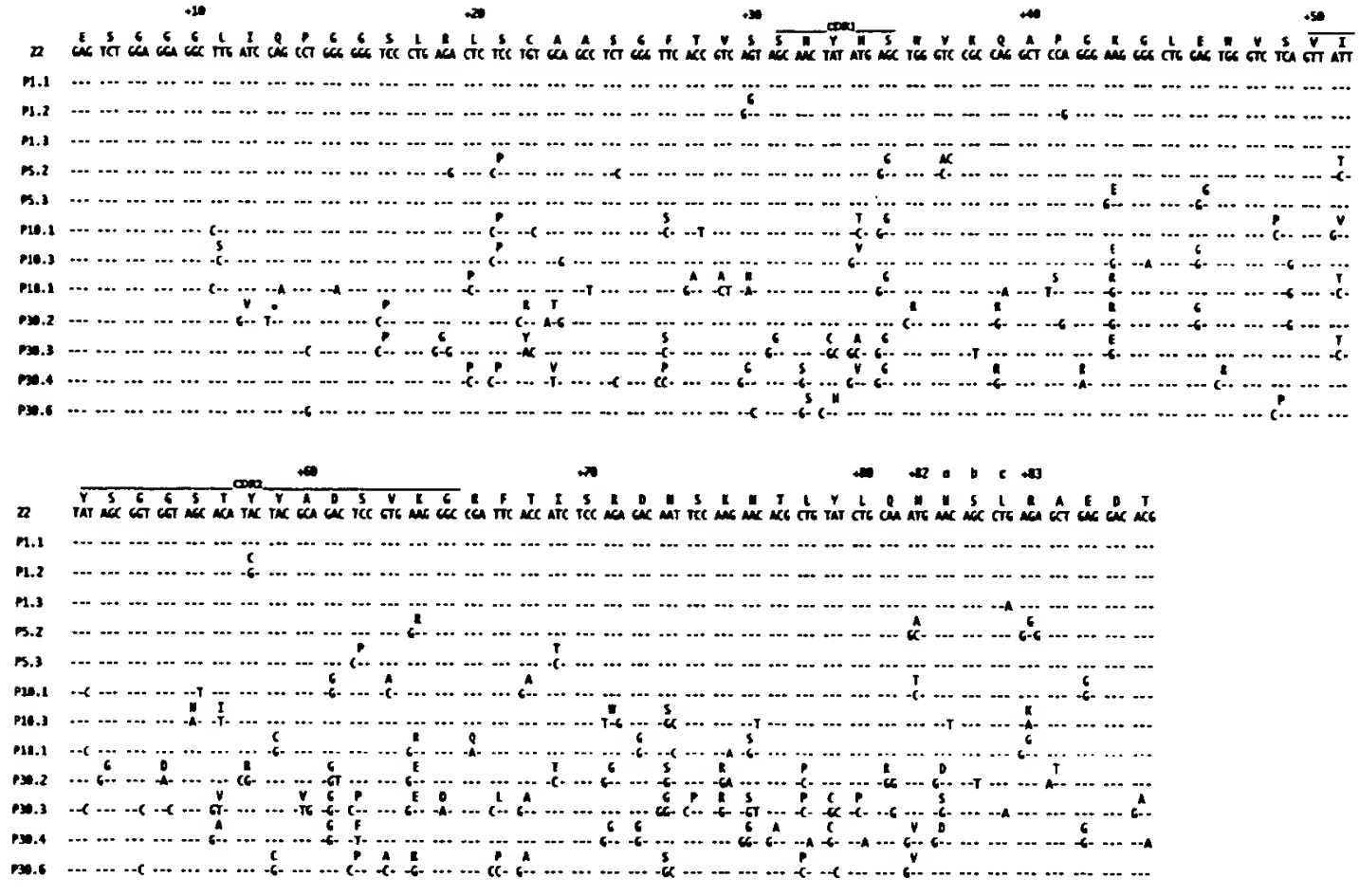
From clones mutagenised with the combination of dPTP and 8-oxodGTP together, the pattern of mutations we observed is shown in Figure 7. All types of transition and transversion mutations which were expected from the combination of the two triphosphate analogues were observed although their respective frequencies were slightly different from those predicted based on the frequencies of dPTP and 8-oxodGTP mutations. The mixture of the two analogues also increased the frequency of additional mutations (1×10^{-3}). No insertions and a single two-nucleotide deletion were found using either analogue over a total of 13,307 bp sequenced.

Amino acid replacements induced by dPTP and 8-oxodGTP in target sequences

We analysed the effect of the four transition mutations induced by dPTP and the two transversion mutations induced by the 8-oxodGTP at the codon level. Figure 8 shows the results of this analysis. The Figure groups amino acids into five classes: glycine, non-polar, polar, positively charged and negatively charged, and shows the codon changes resulting from dPTP mutagenesis (circles), 8-oxodGTP mutagenesis (squares) and their combination (triangles). Codon changes resulting from a single base substitution are shown as full symbols, those resulting from a double substitution are shown as open symbols.

In spite of the clear bias in the mutations induced by dPTP and 8-oxodGTP (Figure 7), the use of these analogues or their combination allowed extensive codon changes to be achieved. The two genes used as model templates contained 51 out of the possible 64 codons (codons not present in either gene are marked with an asterisk in Figure 8). Of the 51

A



B

(continued)

Figure 5(A and B). (*legend on next page*)

C

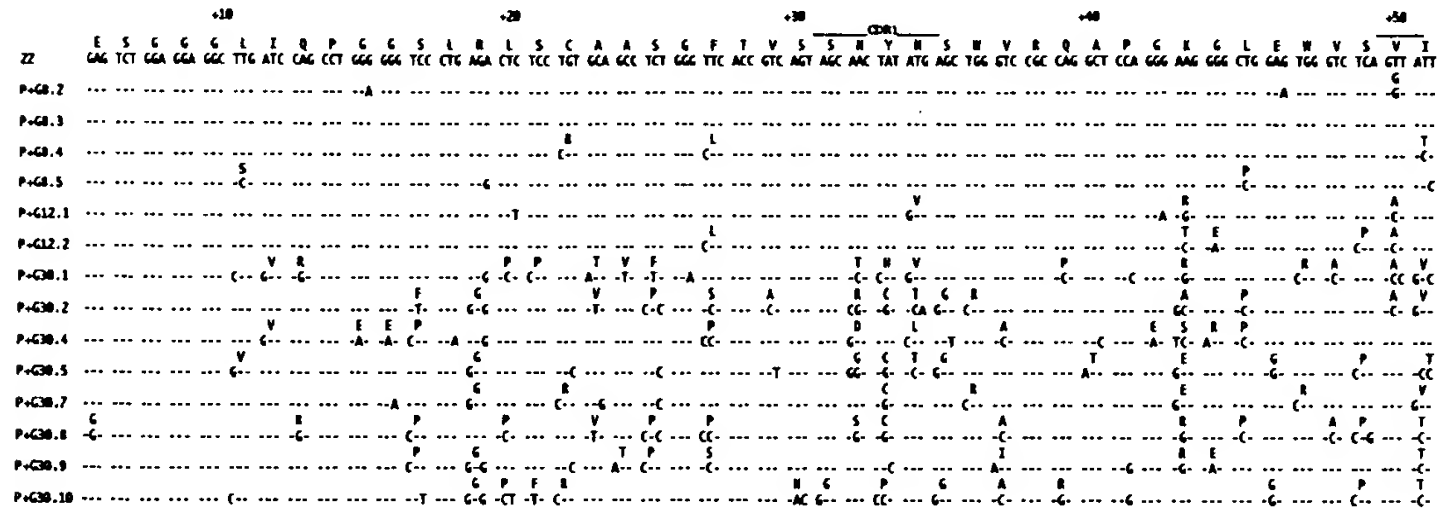


Figure 5. Mutations in the VH segment of the antibody MH22 (Griffiths *et al.*, 1994) and corresponding amino acid substitutions produced by dPTP (A), 8-oxodGTP (B) and the mixture of the two (C) when used in equimolar amount (500 μM) with the four normal dNTPs in a PCR reaction. The number before the dot in the clone designation represents the number of PCR cycles allowed in the presence of the nucleoside triphosphate analogue(s). The second number identifies the clones. The asterisk indicates a stop codon. Amino acid numbers, framework regions (FR) and complementary determining regions (CDR) are defined as described by Kabat *et al.* (1991). See Materials and Methods for complete details of mutagenesis conditions.

codons present, 50 were mutated by dPTP or 8-oxodGTP or by their combination.

Of 224 codon changes which were found at least once in the database, 49 were silent, 66 changed the amino acid to another of the same class, 105 changed the amino acid to one of a different class and four led to termination codons. These results thus demonstrate that a broad spectrum of amino acid substitutions can be generated by dPTP and/or 8-oxodGTP mutagenesis.

Discussion

Comparison of the current procedure with existing methods for random mutagenesis

We have described a procedure that enables rapid and controlled point mutagenesis of DNA sequences *via* mixtures of triphosphate derivatives of base analogues and *Taq* polymerase. The procedure results in a frequency of point mutations, which to the best of our knowledge, is superior to that obtained with any previous method while yielding negligible numbers of nucleotide insertions and deletions.

Two basic strategies have been pursued in the last decade in order to generate random mutations in DNA. In the first, polymerases lacking proof-reading activity, such as AMVs reverse transcriptase, have been used to extend randomly terminated 3'-ends in the presence of a single dNTP (Cunningham & Wells, 1987) or biased dNTP mixtures (Lethovaara *et al.*, 1988). With the availability of *Taq* polymerase, a thermostable DNA polymerase lacking proof-reading activity (Tindall & Kunkel, 1988), several protocols have been developed aimed at harnessing the potential of this enzyme for mutagenesis. The error rate of *Taq* polymerase is of the order of 10^{-4} to 10^{-5} depending on reaction conditions (Eckert & Kunkel, 1990) but can be deliberately increased by altering the concentration of $MgCl_2$ and the ratio of the four dNTPs or by adding $MnCl_2$ (Leung *et al.*, 1989; Cadwell & Joyce, 1992). Although we found that

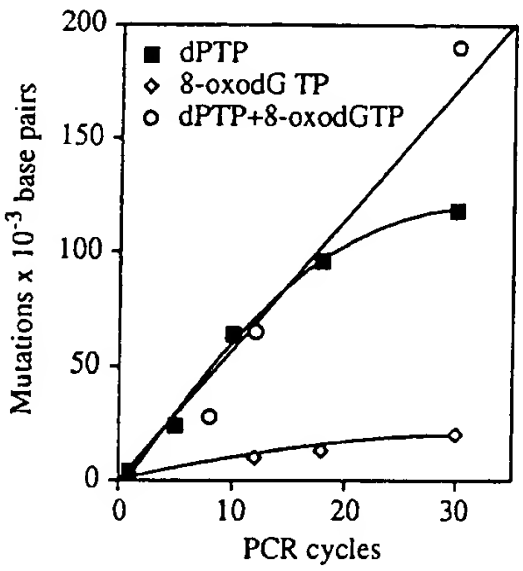


Figure 6. Frequency of mutation *versus* number of cycles of PCR mutagenesis. Reaction conditions are as in Figure 5. The data are derived from two sets of experiments on two separate genes: the VH segment of the MH22 antibody (Griffths *et al.*, 1994) and the Vκ segment of the NQ2/16.2 antibody (Kaartinen *et al.*, 1983).

these protocols yielded a biased pattern of mutations dominated by A → G and T → C transitions (E. Gherardi & C. Milstein, unpublished results), other investigators have succeeded in employing these procedures for the generation and selection of protein variants (Hawkins *et al.*, 1992; Gram *et al.*, 1992).

Over the last few years, however, another approach to the random mutagenesis of DNA has emerged, namely the incorporation into DNA of triphosphate derivatives of pyrimidine and purine base analogues with promiscuous base-pairing potential. This is an approach initially employed by Weissman and colleagues for the generation of site-specific mutations in bacteriophage Qβ using N^4-hydroxy-2'-deoxycytidine triphosphate (Flavell *et al.*, 1974). The naturally-occurring 2'-deoxyinosine has promiscuous base-pairing potential (Ohtsuka *et al.*, 1985) and its triphosphate derivative (dITP), although not a good substrate for *Taq* polymerase (Innis *et al.*, 1988), has also been used in PCR for mutagenesis aided by dNTP pool bias (Ikeda *et al.*, 1992; Spee *et al.*, 1993). However the frequencies of mutation are low and the pattern is dominated by transitions. In addition it is evident that the method is only suitable for mutagenising short segments of DNA due to the poor extension of primers terminating in inosine mismatches (Innis *et al.*, 1988) thereby attenuating amplification of mutagenised fragments.

The analogue dPTP is an excellent substrate for *Taq* polymerase, exhibiting K_m values comparable to those of dCTP and TTP (Table 1). It displays essentially identical kinetic parameters to those of TTP whilst its incorporation opposite template G is about 1/10th as efficient as dCTP incorporation (Table 1). These results are in contrast to those previously obtained during replication of dP within a template using *Taq* polymerase (Kamiya *et al.*, 1994) where only A was incorporated opposite. However, more recent results (F. Hill, D. Loakes & D.M.B., unpublished work) in which a template containing several dP residues was copied by this enzyme, show that both A and G are inserted in a ratio of about 2:1.

We note that N^4-amino-2'-deoxycytidine triphosphate is incorporated by *E. coli* DNA polymerase I half as efficiently as dCTP and one thirtieth as well as TTP (Negishi *et al.*, 1985); it or a close relative is a promising candidate for experiments of the present type, for example to adjust the balance of transition mutations. Clearly the mutagenic pattern of these analogues is directly related to their respective tautomeric constants. In the case of dPTP the mutational pattern which is observed results from the preferential pairing of dP with A, which, by comparison with the catalytic efficiency (V_{max}/K_m) of insertion opposite G (Table 1), is approximately three times as efficient.

8-OxodGTP is known to be incorporated by the Klenow fragment of *E. coli* DNA polymerase I with an efficiency of between (4.4 to 7.2) × 10^{-4} of that of the normal 5'-triphosphates analogues. Although the corresponding values for dPTP have not been determined with this enzyme, it is clear that dPTP is a superior substrate. Since both dPTP and 8-oxodGTP are initially incorporated opposite template adenine residues, their combination produces mutations predominantly originating from dPTP (Figure 7).

Improvements in the current procedure

Three further developments of the approach described herein are envisaged. The first is the replacement of dPTP with a closely related analogue which displays a tautomeric equilibrium that would adjust the balance between all four possible transition mutations. The second concerns the ratio of transition *versus* transversion mutations

Table 2. Numbers and types of mutation produced using dPTP, 8-oxodGTP and their mixture

Mutagenic dNTP	Bases sequenced	Number of point mutations		
		Total	Coding	Silent
dPTP	4093	384	318	66
8-oxodGTP	5463	91	65	16
dPTP and 8-oxodGTP	3751	387	334	53

See Materials and Methods for details of templates used and mutagenesis conditions.

dPTP

from \ to	G	A	T	C
G	-	9.2	0	0
A	46.6	-	0.5	0
T	0.2	0	-	35.5
C	0	0	8.0	-

8-oxo-dGTP

from \ to	G	A	T	C
G	-	1.1	0	0
A	0	-	0	38.8
T	59	0	-	0
C	0	1.1	0	-

dPTP + 8-oxo-dGTP

from \ to	G	A	T	C
G	-	7.8	0	0
A	40.7	-	0.2	6.6
T	1.8	0	-	37.6
C	0.7	0	5.4	-

Figure 7. Pattern of mutations produced by dPTP, 8-oxodGTP and the mixture of the two. The data are obtained from the same experiment used to generate Figure 6 and are expressed as a percentage of the total number of mutations sequenced.

in experiments in which both the dPTP (or related) analogue and 8-oxodGTP are used in combination. It should be possible to obtain comparable numbers of transitions and transversions from mutagenesis reactions in which the concentrations of dPTP (or related) and 8-oxodGTP are adjusted in order to compensate for their different kinetics. Finally, it is clear that six transversion mutations (C → G, T → G, T → A, A → T, A → C and G → C) either are not produced by the dNTP mixture, or else they are produced at very low frequencies (Figure 8). Other analogues therefore, such as O^2-ethylthymidine triphosphate, which induces A → T transversions albeit at a low frequency (Singer *et al.*, 1989), could be used in order to extend the range.

Random mutagenesis *versus* DNA shuffling in protein engineering

While *in vitro* point mutagenesis followed by selection clearly aims to mimic an important aspect of protein evolution, it is clear that nature's strategy of protein engineering equally relies on a variety of other processes such as gene insertions, deletions, duplication and recombination. Procedures are being developed which aim to reproduce these events *in vitro* and harness their potential for protein engineering. In one such procedure, gene fragments obtained by DNase I treatment are reassembled by PCR in a process that promotes random recombination (Stemmer, 1994a). The effectiveness of this approach has been clearly illustrated by its application in the engineering of β-lactamase mutants, one of which, when expressed in *E. coli*, increased the minimum inhibitory concentration (MIC) of a β-lactan antibiotic by 32,000-fold compared to wild-type enzyme (Stemmer, 1994b). It is of interest, however, that the sequence of the improved mutant only contained five point mutations compared to the wild-type enzyme. This shows that Stemmer's protocol leads to an appreciable rate of point mutagenesis and that, at least in the β-lactamase example, such point mutations are entirely responsible for the maturation of the enzyme in the absence of *bona fide* recombination. The results, nevertheless, reinforce the concept that point mutagenesis is a powerful approach for protein engineering *in vitro*.

Mutational load and protein engineering

Previous DNA mutagenesis protocols typically resulted in relatively low mutational rates. The procedure described here, however, can lead to a frequency of nucleotide substitutions approaching one in five after 30 cycles of DNA amplification. This clearly raises the issue of an optimal mutational load for protein engineering.

It seems reasonable to suggest that the lower limit of an efficient random mutagenesis protocol may aim at introducing, on average, one amino acid change per sequence but this may be sub-optimal. While a very large number of simultaneous substitutions would clearly destroy protein stability, studies with several model systems suggest that a relatively small number of amino acids are critical for function and stability.

In T4 lysozyme, for example, substitution of each amino acid (except for the initiator methionine) with 13 different amino acids has shown that more than half the positions tolerated all substitutions (Rennell *et al.*, 1991). Furthermore, out of 2015 mutations, only 173 were seriously deleterious and these were confined to 53 out of 163 positions (Rennell *et al.*, 1991). Studies on the λ repressor also demonstrated that numerous substitutions in the core of the protein are tolerated (Lim & Sauer, 1991).

Although these studies do not address directly the effect of multiple random mutations, they suggest, nevertheless, that these would not invariably result in the loss of protein function, an argument reinforced by the results of studies on somatic hypermutation of antibody genes. The V segments of antibodies isolated in secondary or tertiary responses contain a considerable number of replacement mutations. However in cases in which the role of individual substitutions has been analysed, it appeared that only a few mutations played a role in affinity maturation (for example 3

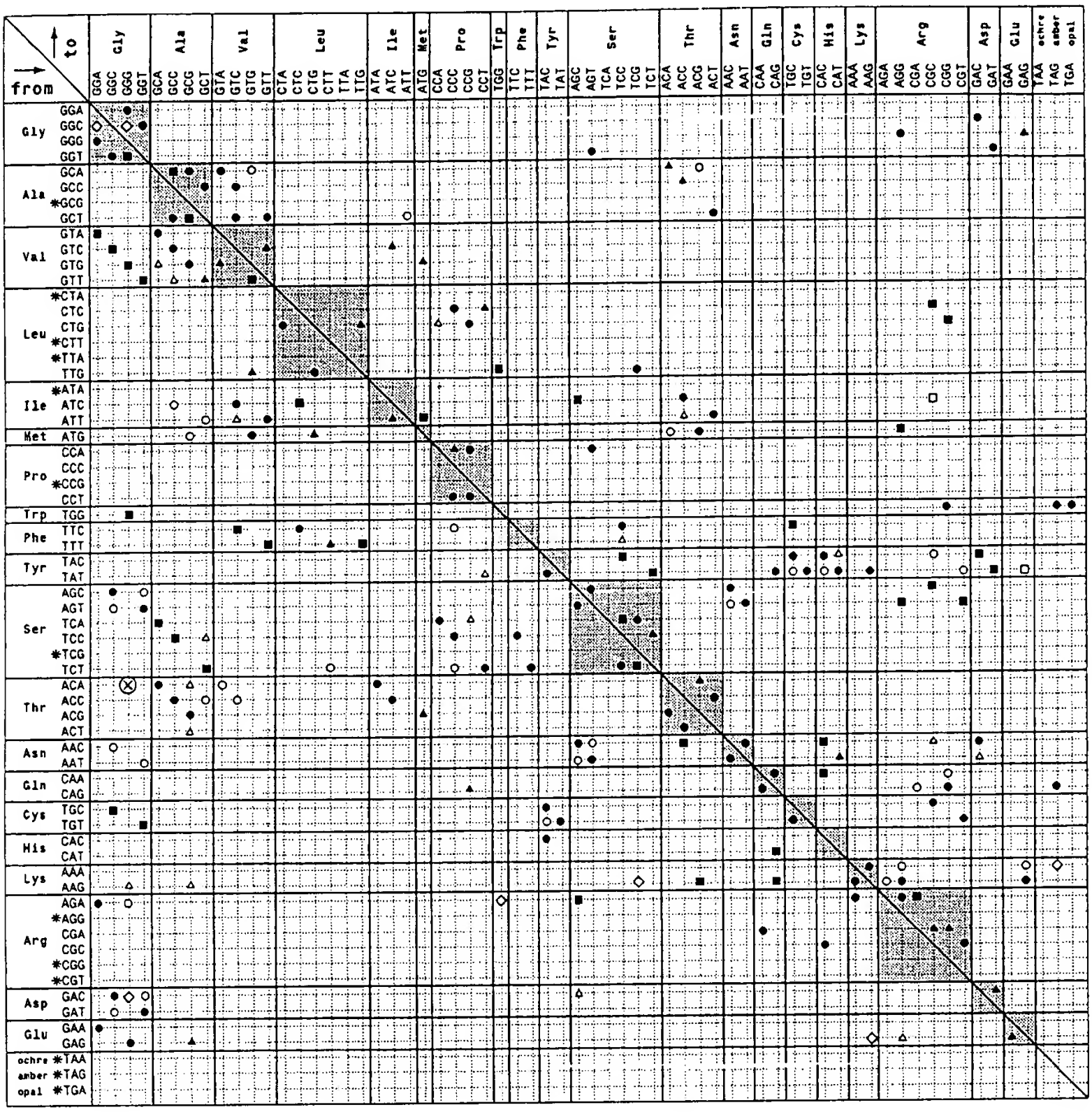


Figure 8. Codon changes produced by dPTP (circles), 8-oxodGTP (squares) and the combination of the two (triangles). Filled symbols indicate a single nucleotide change within a codon, open symbols indicate two nucleotide changes within the same codon. Diamonds indicate the presence of nucleotide changes different from the ones expected (see the text for details). ⊗ Indicates the single case in which three nucleotide changes within a codon were found. Amino acids are grouped in five classes according to their physico-chemical characteristics: glycine, non-polar, polar, positively charged and negatively charged. Asterisks indicate codons which were not present in the two target genes studied.

out of 19 amino acids in the anti-*p*-azophenylarsonate antibody; Sharon, 1990). Yet, the frequency of amino acid substitutions in the VH and VL domains of hypermutated antibodies approaches 1 in 10 (see Berek & Milstein, 1987, for a review). The procedure described here may allow the optimal mutational load for protein engineering to be addressed experimentally since this can be readily controlled (Figure 6) and libraries of protein mutants carrying different numbers of substitutions can be constructed.

Conclusions

Finally, we point out the relationship of the present work to combinatorial oligonucleotide chemistry. In the latter a wide variety of short (n) repertoires, generally of large sequence content (e.g. $\leqslant 4^n$) can be synthesised. Essentially all possible sequence isomers are formed and rounds of selection have to be applied in a variety of formats to identify the sequence of interest. In the present approach a DNA sequence, functional *ab initio*, is

amplified under a variable mutational pressure and the products then cloned. The mutational frequencies that were observed, as related to PCR cycle number were derived from the few tens of randomly picked colonies. Assuming that these mutation frequencies hold over all clones carrying the insert, the number of mutant inserts sequenced represented a very small fraction of the total formed in each amplification process. It is reasonable therefore to propose that very large libraries of mutants can be constructed with the present method and it should be interesting to compare the mutants isolated from such libraries with those obtainable from large synthetic repertoires.

Materials and Methods

6-(2-Deoxy-β-d-ribofuranosyl)-3,4-dihydro-8H-pyrimido -[4,5-c][1,2]oxazin-7-one 5'-triphosphate, triethylammonium salt (dPTP)

The P deoxynucleoside (54 mg, 0.2 mol; Kong Thoo Lin & Brown, 1989) was dried *in vacuo* over P_2O_5 at 80°C overnight, then suspended in dry trimethylphosphate (0.5 ml) under argon. The flask was cooled in an ice-bath whilst phosphoryl chloride (21 µl) was injected with stirring. After stirring in the ice bath for 45 minutes, a vortexed mixture of bis-tributylammonium pyrophosphate (0.5 M in anhydrous DMF, 1.0 ml), tributylamine (0.2 ml) and anhydrous DMF (0.4 ml) was added with rapid stirring in ice, followed after ten minutes by triethylammonium bicarbonate solution (pH 7.5, 0.1 M, 20 ml). After one hour, the sample was purified on Sephadex A25 column (25 mm × 330 mm) at 4°C eluting with a linear gradient of triethylammonium bicarbonate (0.05 to 0.8 M, 3 l). Fractions containing the 5'-triphosphate (0.48 and 0.54 M buffer) were evaporated and purified further by reverse phase HPLC (Waters C18, 7.8 mm × 300 mm) using a linear gradient over 20 minutes of 0 to 4.5% acetonitrile in 0.1 M triethylammonium bicarbonate (pH 7.5) at 2.5 ml/minute. Appropriate fractions were evaporated and residual buffer removed by coevaporation with methanol to afford the pure triphosphate as the *tetrakis*triethylammonium salt (253 A_{260} at pH 7, 0.067 mmol, 34%). δ(^2H_2O) ppm −9.57 (d, γ-P), −10.34 (d, α-P), −22.02 (t, β-P). Approx. HPLC retention time = 18.5 minutes.

8-Oxo-2'-deoxyguanosine 5'-triphosphate, triethylammonium salt (8-oxodGTP)

This was prepared essentially as described (Mo *et al.*, 1985). dGTP (trisodium dihydrate, 58.48 mg, 96 µmol) in 100 mM sodium phosphate (8 ml) containing 30 mM ascorbic acid and 100 mM hydrogen peroxide was incubated at 37°C for four hours in the dark. The product was purified by reverse phase HPLC (Waters C18, 19 mm × 300 mm) using a linear gradient over 30 minutes of 0 to 15% acetonitrile in 0.1 M triethylammonium bicarbonate (pH 7.5) at 7.5 ml/minute. The pure triphosphate was obtained as the *tetrakis*triethylammonium salt as described above. The absorbance spectrum was identical to that described (Mo *et al.*, 1985; Purmal *et al.*, 1994) (12.3 A_{244}, 10.3 A_{293} at pH 7, 5.2 µmol, 5.4%). δ(^2H_2O) ppm −9.68 (d, γ-P), −10.46 (d, α-P), −22.40 (t, β-P). Approx. HPLC retention time = 27.9 minutes, dGTP 26.0 minutes.

Mutagenesis

For mutagenesis experiments, 10 fmol template DNA were amplified using 0.5 µl of AmpliTaq polymerase (5 units/µl, Applied Biosystems) in a 20 µl reaction containing the appropriate sense and antisense primers at 0.5 µM, 2 mM $MgCl_2$, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1 g/l gelatine and dATP, dCTP, dGTP, TTP, dPTP and/or 8-oxodGTP, each at 500 µM. After various numbers of cycles (92°C for one minute, 55°C for 1.5 minute, 72°C for 5 minutes) 1 µl of the amplified material was used in second PCR in which the above conditions were used except that no dPTP or 8-oxodGTP was added to the reaction mixture. The product of the second PCR was restricted with *Bst*EII and *Pst*I and cloned into the M13VHPCR1 vector (Jones *et al.*, 1986). Sequence analysis of single-stranded DNA prepared from phage isolates was performed by the Sanger procedure (Sanger *et al.*, 1977) using Sequenase Version 2 (USB, Cleveland, OH) according to manufacturer's instructions.

Kinetic parameters

Method A: Determination of K_m values for dPTP and TTP using primed M13 DNA

K_m values were determined using primed single-stranded (ss) M13 DNA as template. A mixture of 0.12 µM ss M13mp18 template and 2.5 µM −40 reverse primer was heated at 93°C for three minutes and annealed at 25°C for 20 minutes. 2.5 µl of this stock were diluted in 20 µl reaction mixture containing 2 mM $MgCl_2$, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1 g/l gelatine, dATP, dGTP, dCTP at 12.5 µM each, 1 µl [α-^{32}P]dCTP (3000 Ci/mmol, 10 µCi/µl) and different amounts of TTP or dPTP as indicated. Reactions were initiated by the addition of 0.3 unit *Taq* polymerase to the prewarmed (72°C) reaction mixture. Samples (10 µl) were withdrawn at different time points and reactions stopped by addition of 8 µl of 20 µM EDTA in 95% formamide. To remove unincorporated label the reaction product was precipitated three times in ethanol/sodium acetate and DNA synthesis was measured as percentage of [^{32}P]dCTP incorporation. K_m values were determined by the direct linear plot method (Eisenthal & Cornish-Bowden, 1974).

Method B: Determination of kinetic parameters for insertion of dPTP and TTP opposite template A, and dPTP and dCTP opposite template G

Specific kinetic parameters for the insertion of dP opposite A or G, and T opposite A and C opposite G were determined essentially as described (Boosalis *et al.*, 1987) using a 23 base primer and one of two templates (Figure 4). Both the primer and template oligonucleotides were purified DMT-on by reverse phase HPLC (Waters C18, 7.8 mm × 300 mm) with a linear gradient of 2.5 to 50% acetonitrile in 0.1 M triethylammonium bicarbonate (pH 7.5) at 2.5 µl/minute. Appropriate fractions were evaporated and the oligonucleotides detritylated with 80% acetic acid for 30 minutes, then repurified by RP-HPLC as described above. Anion exchange chromatography (Hichrom, SAX 10) using an isocratic gradient of 21 mM potassium phosphate (pH 6.3) in 60% formamide solution for two minutes then a linear gradient of 21 to 300 mM in 33 minutes at 2.5 ml/minute followed by dialysis gave the pure oligonucleotides.

Primer end labelling. The primer was 5'-end labelled with ^{32}P in a reaction buffer (30 µl) of 70 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$ and 5 mM DTT, containing 6.7 µM primer and 2.2 µM [γ-^{32}P]ATP (3000 Ci/mmol), and five units of T4 polynucleotide kinase. The solution was incubated at 37°C for 30 minutes then cold ATP (10 mM, 5 µl) added and after a further 10 minutes, the reaction was terminated by heating at 65°C for 15 minutes. The labelled oligonucleotide was purified by 20% PAGE and eluted in 3 M sodium acetate solution (pH 5.5). The primer was then ethanol precipitated, redissolved in water and desalted on a Sephadex G-25 (NAP-10) column.

Primer-template annealing. The ^{32}P-labelled primer was annealed to the respective template in a 500 µl of 1 × buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl) containing 78.9 nM template and 86.8 nM primer at 95°C for five minutes, then cooled slowly to room temperature.

Determination of kinetic parameters

Taq DNA polymerase (Perkin Elmer Native *Taq* five units/µl) was diluted with buffer containing 20 mM Tris-HCl (pH 8.0), 100 mM KCl 0.1 mM EDTA, 1 mM DTT and 50% glycerol to 0.2 unit/µl, then diluted to 0.0066 unit/µl in 1 × buffer. Solution A contained diluted *Taq* polymerase (30 µl) and annealed primer-template solution (225 µl) in 1 × buffer. Solution B contained 355 µM dGTP, 5.33 µM $MgCl_2$ and dPTP, dCTP or TTP at an appropriate concentration. Solution A (25.5 µl) containing *Taq* polymerase and annealed primer-template and solution B (25 µl) containing dNTPs/$MgCl_2$ were preincubated at 55°C for four minutes and reactions were initiated by the addition of 10 µl of B to solution A. Final reaction mixtures contained 1.5 mM $MgCl_2$, 100 mM dGTP, 50 nM annealed primer-template and concentrations of dPTP of between 0.5 and 100 mM, and dCTP and TTP between 0.5 and 30 mM. For dPTP concentrations between 0.5 and 100 mM or dCTP and TTP concentrations between 0.5 and 30 mM initial velocities were determined up to ten minutes as described (Boosalis *et al.*, 1987) by terminating 8 µl aliquots of reaction mixture with 12 µl of 95% formamide/10 mM EDTA containing 0.5% (w/v) bromophenol blue. Samples were denatured by heating at 95°C for three minutes, then 10 µl were electrophoresed at 38 W for three hours on a 16% (w/v) denaturing polyacrylamide gel and analysed using a Molecular Dynamics PhosphorImager. Bands were quantified using the GelTrak program (Smith & Thomas, 1990). Initial velocities were constant with these substrate concentrations for up to ten minutes, therefore samples at t = four minutes (dPTP) or three minutes (dCTP and TTP) using varying dNTP concentrations were chosen for the determination of kinetic parameters.

Acknowledgements

The authors are very grateful to Drs C. Milstein and P. J. G. Butler for valuable discussions. The work was supported by the Human Genome Mapping Project (MRC) and by the BIOTECH programme (CEC). E.G. also has a part-time teaching position in immunology at the University of Pavia (Italy).

References

Abdul-Mashi, M. T. & Bessman, M. J. (1986). Biochemical studies on the mutagen 6-*N*-hydroxylaminopurine. *J. Biol. Chem.* **261**, 2020–2026.

Anand N. N., Brown, D. M. & Salisbury, S. A. (1987). The stability of oligodeoxynucleotide duplexes containing degenerate bases. *Nucl. Acids Res.* **15**, 8167–8176.

Berek, C. & Milstein, C. (1987). Mutation drift and repertoire shift in the maturation of the immune response. *Immunol. Rev.* **96**, 23–41.

Boosalis, M. S., Petruska, J. & Goodman, M. F. (1987). DNA polymerase insertion fidelity. *J. Biol. Chem.* **262**, 14689–14696.

Brown, D. M., Hewlins, M. J. E. & Schell, P. (1968). The tautomeric state of *N*(4)-hydroxy- and *N*(4)-amino-cytosine derivatives. *J. Chem. Soc. C* 1925–1929.

Cadwell, R. C. & Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *PCR Methods Applic.* **2**, 28–33.

Cheng, K. C., Cahill, D. S., Kasai, H., Nishimuta, S. & Loeb, L. A. (1992). 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G to T and A to C substitutions. *J. Biol. Chem.* **267**, 166–172.

Cunningham, B. C. & Wells, J. A. (1987). Improvement in the alkaline stability of subtilisin using an efficient random mutagenesis and screening procedure. *Protein Eng.* **1**, 319–325.

Dower, W. J., Miller, J. F. & Ragsdale, C. W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucl. Acids Res.* **16**, 6127–6145.

Eadie, J. S., Conrad, M., Toorchen, D. & Topal, M. D. (1984). Mechanism of mutagenesis by O^6-methylguanine. *Nature*, **308**, 201–203.

Eckert, K. A. & Kunkel, T. A. (1990). High fidelity DNA synthesis by the *Thermus acquaticus* DNA polymerase. *Nucl. Acids Res.* **18**, 3739–3744.

Eisenthal, R. & Cornish-Bowden, A. (1974). The direct linear plot. A new graphical procedure for estimating enzyme kinetic parameters. *Biochem. J.* **139**, 715–720.

Estell, D. A., Graycar, T. P. & Wells, J. A. (1985). Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *J. Biol. Chem.* **260**, 6518–6521.

Flavell, R. A., Sabo, D., Bandle, E. & Weissman, C. (1974). Site-directed mutagenesis: generation of an extracistronic mutation in bacteriophage Qβ. *J. Mol. Biol.* **89**, 255–272.

Gillam, S. & Smith, M. (1979a). Site-specific mutagenesis using synthetic oligonucleotide primers: I. Optimum conditions and minimum oligodeoxy ribonucleotide length. *Gene*, **8**, 81–97.

Gillam, S. & Smith, M. (1979b). Site-specific mutagenesis using synthetic oligonucleotide primers: II. In vitro selection of mutant DNA. *Gene*, **8**, 99–106.

Gram, H., Marconi, L. A., Barbas III, C. F., Collet, T. A., Lerner, R. A. & Kang, A. S. (1992). *In vitro* selection and affinity maturation of antibodies from naive combinatorial immunoglobulin libraries. *Proc. Natl Acad. Sci. USA*, **89**, 3576–3580.

Griffiths, A. D., Williams, S. C. Hartley, O., Tomlinson, I. M., Waterhouse, P., Crosby, W. L., Kontermann, R. E., Jones, P. T., Low, N. M., Allison, T. J., Prospero, T. D., Hoogenboom, H. R., Nissim, A., Cox, J. P. L., Harrison, J. L., Zaccolo, M., Gherardi, E. & Winter, G. (1994). Isolation of high-affinity human antibodies directly from large synthetic repertoires. *EMBO J.* **13**, 3245–3260.

Grossberger, D. & Clough, W. (1981). Incorporation into

DNA of the base analogue 2-aminopurine by Epstein Barr virus induced DNA polymerase in vivo and in vitro. *Proc. Natl Acad. Sci. USA*, **78**, 7271–7275.

Hawkins, R. E., Russell, S. J. & Winter, G. (1992). Selection of phage antibodies by binding affinity. Mimicking affinity maturation. *J. Mol. Biol.* **226**, 889–896.

Hutchinson, C. A., Phillips, S., Edgell, M. N. Gillam, S. Jahnke, P. & Smith, M. (1978). Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.* **253**, 6551–6560.

Ikeda, M., Hamano, K. & Shibata, T. (1992). Epitope mapping of anti-recA protein IgGs by region specified polymerase chain reaction mutagenesis. *J. Biol. Chem.* **267**, 6291–6296.

Innis, M. A., Myambo, K. B., Gelfand, D. H. & Brow M. A. D. (1988). DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc. Natl Acad. Sci. USA*, **85**, 9436–9440.

Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G. (1986). Replacing the complementary-determining regions in human antibody with those from a mouse. *Nature*, **321**, 522–525.

Kabat, E. A., Wu, T. T., Perry, H. M. Gottesman, K., S. & Foeller, C. (1991). *Sequences of Proteins of Immunological Interest*, 5th edit., U. S. Department of Health and Human Services, Public Health Services National Institute of Health.

Kamiya, H., Murata-Kamiya, N., Kong Thoo Lin, P., Brown, D. M. & Ohtsuka, E. (1994). Nucleotide incorporation opposite degenerate bases by *Taq* polymerase. *Nucleosides Nucleotides*, **13**, 1483–1492.

Kaartinen, M., Griffiths, G. M., Markham, A. F. & Milstein, C. (1983). mRNA sequences define an unusually restricted IgG response to 2-phenyl-oxazolone and its early diversification. *Nature*, **304**, 320–325.

Komiyama, N. H., Miyazaki, G., Tame, J. & Nagai, K. (1995). Transplanting a unique, allosteric effect from crocodile into human haemoglobin. *Nature*, **373**, 244–246.

Kong, H., Kucera, R. B. & Jack, W. E. (1993). Characterisation of a DNA polymerase from the Hyperthermophile Archea *Thermococcus litoralis*. *J. Biol. Chem.* **268**, 1965–1975.

Kong Thoo Lin, P. & Brown, D. M. (1989). Synthesis and duplex stability of oligonucleotides containing cytosine-thymine analogues. *Nucl. Acids Res.* **17**, 10373–10383.

Kouchakdjian, M., Bodepudi, V., Shibutani, S., Eisenberg, M., Johnson, F., Grollman, A. P. & Patel, D. J. (1991). NMR structural studies of the ionizing radiation adduct 7-hydro-8-oxodeoxyguanosine (8-oxo-*7H*-dG) opposite deoxyadenosine in a DNA duplex. 8-oxo-*7H*-dG(*syn*).dA(*anti*) alignment at lesion site. *Biochemistry*, **30**, 1403–1412.

Lehtovaara, P. M., Koivula, A. K., Bamford, J. & Knowles, J. K. C. (1988). A new method for random mutagenesis of complete genes: enzymatic generation of mutant libraries *in vitro*. *Protein Eng.* **2**, 63–68.

Leung, D. W., Chen, E. & Goeddel, D. V. (1989). A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Techniques*, **1**, 11–15.

Lim, W. A. & Sauer, R. T. (1991). The role of internal packing interactions in determining the structure and stability of a protein. *J. Mol. Biol.* **219**, 359–376.

Looker, D., Abbott-Brown, D., Cozart, P., Durfee, S., Hoffman, S., Mathews, A. J., Miller-Roerich, J., Shoemaker, S., Trimble, T., Fermi, G., Komiyama. N. H., Nagai, K. & Stetler, G. L. (1992). A human recombinant haemoglobin designed for use as a blood substitute. *Nature*, **356**, 258–260.

Ludwig, J. (1981). A new route to nucleoside 5'-triphosphates (1981). *Acta Biochim. Biophys. Acad. Sci. Hung.* **16**, 131–133.

Maki, A. & Sekiguchi, M. (1992). MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. *Nature*, **355**, 273–275.

Matsumoto, K., Yashiki, T., Bessho, T., Negishi, K. & Hayatsu, H. (1982). Mutations induced by *in vitro* incorporation of N^4-aminodeoxycytidine 5'-triphosphate into DNA of phage M13mp2. *Mutat. Res.* **268**, 59–64.

Mo, J. Y., Maki, H. & Sekiguchi, M. (1992). Hydrolytic elimination of a mutagenic nucleotide, 8-oxodGTP, by human 18-kilodalton protein: sanitization of the nucleotide pool. *Proc. Natl Acad. Sci. USA*, **89**, 11021–11025.

Moore, M. H., van Meervelt, L., Salisbury, S. A., Kong Thoo Lin, P. & Brown, D. M. (1995). Direct observation of two base-pairing modes of a cytosine-thymine analogue with guanine in a DNA *Z*-form duplex; Significance for base analogue mutagenesis. *J. Mol. Biol.* **251**, 665–673.

Mott, J. E., van Arsdell, J. & Platt, T. (1984). Targeted mutagenesis *in vitro: lac* repressor mutations generated using AMV reverse transcriptase and dBrUTP. *Nucl. Acids Res.* **12**, 4139–4152.

Morozov, Y. V, Savin, F. A., Chechov, V. O., Budowsky, E. I. & Yakovlev, D. Y. (1982). Photochemistry of N6-methoxyadenosine and of N4-hydroxycytidine and its methyl derivatives. *J. Photochem.* **20**, 229–252.

Müller, W., Weber, H., Meyer, F. & Weissman, C. (1978). Site-directed mutagenesis in DNA: Generation of point mutations in cloned β globin complementary DNA at the positions corresponding to amino acids 121 to 123. *J. Mol. Biol.* **124**, 343–358.

Nedderman, A. N. R., Stone, M. J., Williams, D. H., Kong Thoo Lin, P. & Brown, D. M. (1993). Molecular basis for methoxyamine initiated mutagenesis: ^{1}H nuclear magnetic resonance studies of oli gonucleotide duplexes containing base-modified cytosine residues. *J. Mol. Biol.* **230**, 1068–1076.

Negishi, K., Takahashi, M., Yamashita, Y., Nishizawa M. & Hayatsu, H. (1985). Mutagenesis by N4-aminocytidine: induction of AT to GC transition and its molecular mechanism. *Biochemistry*, **24**, 7273–7278.

Neuberger, M. S., Williams, G. T. & Fox, R. O. (1984). Recombinant antibodies possessing novel effector functions. *Nature*, **312**, 604–608.

Oda, Y., Uesugi, H., Ikehara, M., Nishimura, S., Kawase, Y., Ishikawa, H., Inoue, H. & Ohtsuka, E. (1991). NMR studies of a DNA containing 8-hydroxydeoxyguanosine. *Nucl. Acids Res.* **19**, 1407–1412.

Ohtsuka, E., Shigeru, M., Ikehara, M., Takahashi, Y. & Matsubara, K. (1985). An alternative approach to deoxyoligonucleotides as hybridisation probes by insertion of deoxyinosine at ambiguous positions. *J. Biol. Chem.* **260**, 2605–2608.

Pavlov, Y. I., Minnik, D. T., Izuta, S. & Kunkel, T. A. (1994). DNA replication fidelity with 8-oxodeoxyguanosine triphosphate. *Biochemistry*, **33**, 4695–4701.

Purmal, A. A., Wah Kow Y. & Wallace S. S. (1994). 5-Hydroxypyrimidine deoxynucleoside triphosphates are more efficiently incorporated into DNA by exonuclease-free Klenow fragment than 8-oxopurine

deoxynucleoside triphosphates. *Nucl. Acids Res.* **22**, 3930–3935.

Reeves, S. T. & Beattie, K. L. (1985). Base-pairing properties of N4-methoxydeoxycytidine 5'-triphosphate during DNA synthesis on natural templates, catalysed by DNA polymerase I of *Escherichia coli*. *Biochemistry*, **24**, 2262–2268.

Rennell, D., Bouvier, S. E., Hardy, L. W. & Poteete, A. R. (1991) Systematic mutation of bacteriophage T4 lysozyme. *J. Mol. Biol.* **222**, 67–87.

Sanger, F., Niklen, S. & Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5466.

Sharon, J. (1990). Structural correlates of high antibody affinity: three engineered aminoacid substitutions can increase the affinity of an anti p-axophenylarsonate antibody 200 fold. *Proc. Natl Acad. Sci. USA*, **87**, 4814–4817.

Shibutani, S., Takeshita, M. & Grollman, A. P. (1991) Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature*, **349**, 431–434.

Shugar, D., Huber, C. P. & Birnbaum, G. I. (1976). Mechanism of hydroxylamine mutagenesis. Crystal structure and conformation of 1,5-dimethyl-N4-hydroxycytosine. *Biochim. Biophys. Acta*, **447**, 274–284.

Singer, B., Fraenkel-Conrat, H., Abbott, L. G. & Spengler, S. J. (1984). N^4-Methoxydeoxycytidine triphosphate is in the imino tautomeric form and substitutes for deoxythymidine triphosphate in primed poly d[A-T] synthesis with *E. coli* DNA polymerase I. *Nucl. Acids Res.* **12**, 4609–4619.

Singer, B., Chavez, F. & Spengler, S. J. (1986). O^4-Methyl-, O^4-ethyl-, and O^4-isopropylthymidine 5'-triphosphates as analogues of thymidine 5'-triphosphate: kinetics of incorporation by *Escherichia coli* DNA polymerase I. *Biochemistry*, **25**, 1201–1205.

Singer, B., Chavez, F., Spengler, S. J., Kusmierek, J. T., Mendelman, L. & Goodman, M. F. (1989). Comparison of polymerase insertion kinetics of O^2-alkyldeoxythymidine triphosphates and O^4-methyldeoxythymidine triphosphate. *Biochemistry*, **28**, 1478–1483.

Smith G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, **228**, 1315–1317.

Smith, J. M. & Thomas, D. J. (1990). Quantitative analysis of one-dimensional gel electrophoresis profiles. *CABIOS*, **6**, 93–99.

Snow, E. T., Foote, R. S. & Mitra, S. (1984). Kinetics of incorporation of O^6-methyldeoxyguanosine monophosphate during in vitro DNA synthesis. *Biochemistry*, **23**, 4289–4294.

Spee, J. H., de Vos, W. M. & Kuipers, O. P. (1993). Efficient random mutagenesis method with adjustable mutation frequency by use of PCR and dITP. *Nucl. Acids Res.* **21**, 777–778.

Stemmer, W. P. C. (1994a). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl Acad. Sci. USA*, **91**, 10747–10751.

Stemmer, W. P. C. (1994b). Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, **370**, 389–391.

Stone, M. J., Nedderman, A. N. R., Kong Thoo Lin, P., Brown, D. M. & Williams, D. H. (1991). Molecular basis for methoxyamine initiated mutagenesis ^{1}H nuclear magnetic resonance studies of base-modified oligodeoxynucleotides. *J. Mol. Biol.* **222**, 711–723.

Tindall, K. R. & Kunkel, T. A. (1988). Fidelity of DNA synthesis by the *Thermus acquaticus* DNA polymerase. *Biochemistry*, **27**, 6008–6013.

Winter, G., Fersht, A. R., Wilkinson, A. J., Zoller, M. & Smith, M. (1982). Redesigning enzyme structure by site-directed mutagenesis: tyrosyl tRNA synthetase and ATP binding. *Nature*, **299**, 756–758.

Edited by A. R. Fersht

(Received 12 September 1995; accepted 2 November 1995)

Nucleic Acids Research, 1998, Vol. 26, No. 9 2105–2111

Synthesis and RNA polymerase incorporation of the degenerate ribonucleotide analogue rPTP

K. Moriyama, K. Negishi, M. S. J. Briggs[1], C. L. Smith[1], F. Hill[2], M. J. Churcher[2], D. M. Brown[2] and D. Loakes[3,*]

Gene Research Centre, Okayama University, Tsushima, Okayama 700, Japan, [1]Nycomed Amersham plc, Amersham Laboratories, White Lion Road, Amersham, Buckinghamshire HP7 9LL, UK and [2]Laboratory of Molecular Biology and [3]Medical Research Council, Centre for Protein Engineering, Hills Road, Cambridge CB2 2QH, UK

Received January 22, 1998; Revised and Accepted March 19, 1998

ABSTRACT

The synthesis and enzymatic incorporation into RNA of the hydrogen bond degenerate nucleoside analogue 6-(β-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido[4,5-c]-[1,2]oxazin-7-one (P) is described. The 5′-triphosphate of this analogue is readily incorporated by T3, T7 and SP6 RNA polymerases into RNA transcripts, being best incorporated in place of UTP, but also in place of CTP. When all the uridine residues in an HIV-1 TAR RNA transcript are replaced by P the transcript has similar characteristics to the wild-type TAR RNA, as demonstrated by similar melting temperatures and CD spectra. The P-substituted TAR transcript binds to the Tat peptide ADP-1 with only 4-fold lowered efficiency compared with wild-type TAR.

INTRODUCTION

Nucleoside bases differing from the normal purines, adenine and guanine, and the pyrimidines, thymine (uracil) and cytidine, are uncommon in DNA but relatively abundant in RNA, and in particular in transfer RNAs. The vast majority are the result of post-transcriptional modifications of the nucleic acids by specific enzymes. Although a specific role can seldom be assigned to these modifications, their importance can be inferred from their remarkable phylogenetic conservation. Analogues of the natural bases which could be incorporated enzymatically *in vitro* could provide useful functional alterations to synthetic RNA transcripts. Early mutagenesis studies have demonstrated that N^4-hydroxycytidine triphosphate (1, Fig. 1) is efficiently incorporated by the DNA-dependent RNA polymerase from *Micrococcus luteus* (1): the phage Qβ, T2 and *Escherichia coli* Pol II polymerases also incorporate it into RNA or DNA (2–4). The addition of the electronegative element to the N^4-amino group alters the tautomeric ratio of the base; the alternative tautomers can base-pair with either adenine or guanine. The tautomeric constant (K_T) for 1-methyl-N^4-hydroxycytosine has been measured giving a ratio of 10:1 in favour of the oximino-form in water (5), and thus correlates qualitatively with the analogue being recognised as either of the natural pyrimidines.

The N^4-hydroxyl group can adopt either a *syn* or an *anti* conformation; the preferred *syn* form (6), however, interferes with hydrogen bonding of Watson–Crick base-pairs, and this is more evident in the case of N^4-methoxy-derivatives (6). To constrain the hydroxyl group in an *anti* conformation we have previously synthesised a ribonucleoside containing a 5-membered second ring (2, Fig. 1) (7), but this compound proved too unstable to allow conversion either to its phosphoramidite monomer or its 5′-triphosphate. Strain in the 5-membered ring resulted in its cleavage during further reactions. The 2′-deoxynucleoside (dP) containing a 6-membered second ring proved to be stable. The properties of dP and its 5′-triphosphate have been intensively studied (8–11). The ribonucleoside 6-(β-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido[4,5-c][1,2]oxazin-7-one (rP) (3) has now been synthesised, by a route modified from that described for the deoxynucleoside, and converted to its 5′-triphosphate. Incorporation of the triphosphate by three widely used RNA polymerases, those of the bacteriophages T3, T7 and SP6 has been examined. The properties of HIV-1 TAR RNA transcripts synthesised by T3 RNA polymerase using rPTP in place of UTP or CTP have also been investigated.

MATERIALS AND METHODS

General methods

^{1}H NMR spectra were obtained on Bruker WM-250 and DRX 300, and ^{31}P NMR spectra on a Bruker WM-250 spectrometer. NMR spectra were obtained in d^6-DMSO. ^{31}P NMR spectra are referenced to phosphoric acid. Mass spectra were recorded on a Hewlett-Packard G205A Maldi-TOF spectrometer with positive polarity, in a matrix of α-cyano-4-hydroxycinnamic acid in MeCN:H_2O (1:1) with 3% trifluoroacetic acid. UV spectra were recorded on a Perkin Elmer Lambda 2 spectrophotometer fitted with a Peltier cell and samples were dissolved in 1% aqueous methanol. TLC was carried out on pre-coated F_{254} silica plates and column chromatography with Merck kieselgel 60. Unless otherwise stated reactions were worked up as follows: after

*To whom correspondence should be addressed. Tel: +44 1223 248011; Fax: +44 1223 412178; Email: dml@mrc-lmb.cam.ac.uk

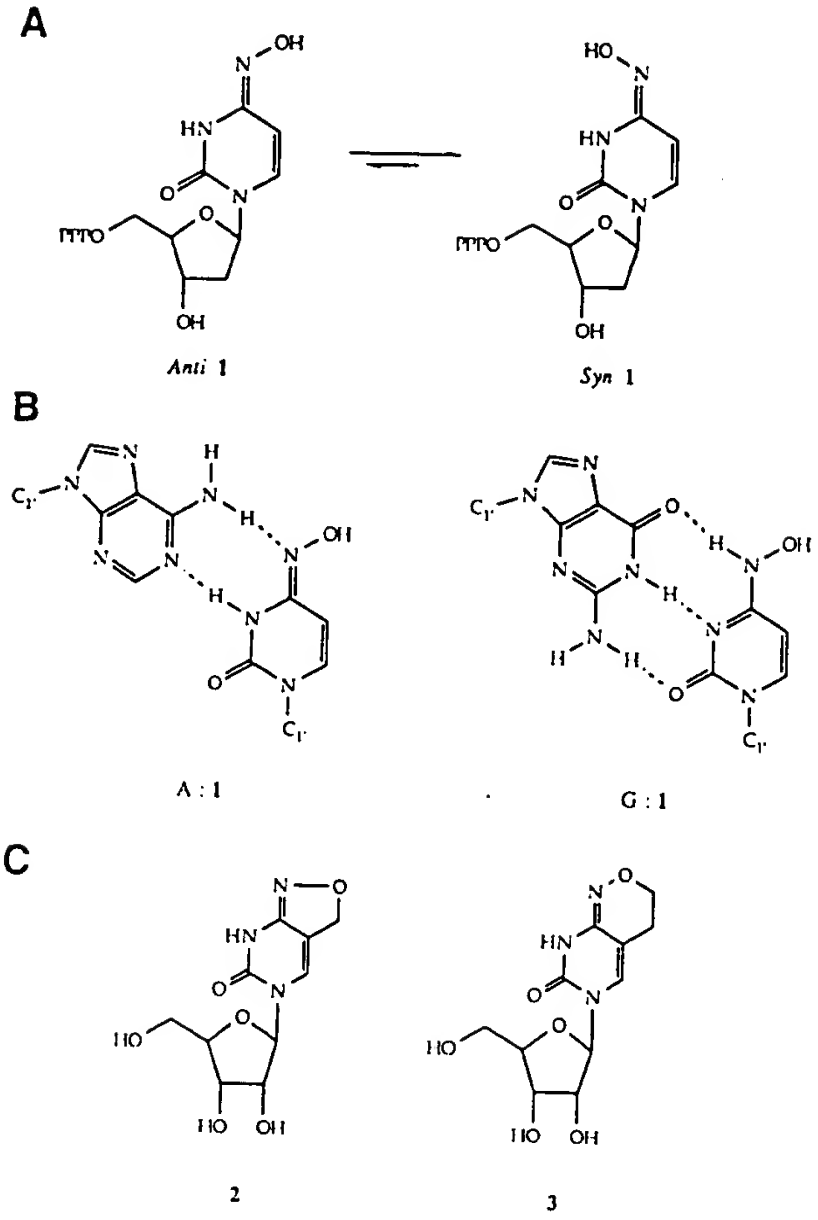


Figure 1. (A) *Syn–anti* conformations of N^4-hydroxycytidine-5′-triphosphate. (B) Base-pairing of N^4-hydroxycytidine with adenosine and guanosine. (C) Structures of degenerate bicyclic ribonucleosides.

removal of the solvent, the product was dissolved in chloroform and washed with aqueous sodium bicarbonate solution. The combined organic fractions were dried over sodium sulphate and evaporated.

Synthesis

5-Allyluridine was prepared according to literature procedures (12). This was then converted to the 2′,3′,5′-tris-(tert-butyldimethylsilyl) derivative (**4**) as previously described (13).

2′,3′,5′-Tri-(tert-butyldimethylsilyl)-5-(2,3-dihydroxypropyl)-uridine (5). To a solution of 2′,3′,5′-tri-(tert-butyldimethylsilyl)-5-allyluridine (**4**) (9.5 g, 15 mmol) in acetone (250 ml) was added *N*-methylmorpholine-*N*-oxide (5.4 g, 46 mmol) followed by potassium osmate dihydrate (70 mg) in water (10 ml) over 5 min. The solution was then stirred at room temperature overnight, and the solvent was removed. The product was worked up as described and then chromatographed ($CHCl_3$/5% MeOH) to give a white foam. Yield 9.84 g, 98%. ^{1}H NMR δ (p.p.m.) –0.11–0.09 (18H, m, 6× $SiCH_3$), 0.79–0.89 (27H, m, 3× $C(CH_3)_3$), 3.20–3.31 (2H, m, CH_2CH), 3.52–3.55 (3H, m, $CHCH_2$OH), 3.68–3.77 (2H, m, H5′, H5″), 3.91–3.92 (1H, m, H2′), 4.03 (1H, t, H3′), 4.19–4.21 (1H, m, H4′), 4.49 (1H, t, OH), 4.56 (1H, d, OH), 5.85 (1H, t, J = 7.4 Hz, H1′), 7.36 (1H, d, J = 5.3 Hz, H6), 11.40 (1H, s, NH). UV λ_{max} 268, 208; λ_{min} 230. M/z 684.306 $(M+Na)^+$, 700.809 $(M+K)^+$.

2′,3′,5′-Tri-(tert-butyldimethylsilyl)-5-(2-hydroxyethyl)-uridine (7). The above diol (**5**) (8.3 g, 12.6 mmol) was dissolved in dioxane (125 ml, 10 ml/mmol) and water (25 ml, 2 ml/mmol) added. To this was then added a solution of sodium periodate (8.0 g, 37 mmol) in water (25 ml, 2 ml/mmol) and the solution stirred at room temperature for 2.5 h. The solution was concentrated then worked up as usual to give the crude aldehyde, **6**. This was dissolved in THF (200 ml) and sodium borohydride (0.5 g, 13 mmol) added followed by water (1 ml) and the solution stirred at room temperature for 1 h. The reaction was quenched with acetic acid, the solution evaporated and the product worked up as usual and chromatographed ($CHCl_3$/3% MeOH) to give an off-white foam. Yield 4.42 g, 56%, remainder (3.06 g) unreacted diol **5**. ^{1}H NMR δ (p.p.m.) –0.10–0.10 (18H, m, 6× $SiCH_3$), 0.79–0.90 (27H, m, 3× $C(CH_3)_3$), 2.33–2.37 (2H, m, CH_2CH_2OH), 3.36–3.43 (2H, m, CH_2CH_2OH), 3.69–3.79 (2H, m, H5′, H5″), 3.91 (1H, br s, H2′), 4.02–4.04 (1H, m, H3′), 4.18–4.22 (1H, m, H4′), 4.61 (1H, t, OH), 5.86 (1H, d, J = 7.1 Hz, H1′), 7.40 (1H, s, H6), 11.40 (1H, s, NH). UV λ_{max} 267, 209; λ_{min} 230. M/z 654.487 $(M+Na)^+$, 671.172 $(M+K)^+$.

2′,3′,5′-Tri-(tert-butyldimethylsilyl)-5-(2-phthalimidooxyethyl)-uridine (8). To a solution of the alcohol (**7**) (3.1 g, 5 mmol) in THF (50 ml) was added triphenyl phosphine (2.6 g, 1 mmol), *N*-hydroxyphthalimide (1.6 g, 1 mmol) and then diisopropylazodicarboxylate (DIAD) (2 g, 1 mmol) and the solution stirred at room temperature overnight. The solution was then evaporated, worked up as described and then chromatographed (twice, $CHCl_3$/1% MeOH) to give a pale yellow foam. Yield 3.65 g, 96%. ^{1}H NMR δ (p.p.m.) –0.11–0.10 (18H, m, 6× $SiCH_3$), 0.58–0.91 (27H, m, 3× $C(CH_3)_3$), 2.61–2.71 (2H, m, CH_2CH_2ON), 3.60–3.69 (2H, m, CH_2CH_2ON), 3.72–3.90 (3H, m, H2′, H5′, H5″), 4.04 (1H, br s, H3′), 4.14–4.27 (1H, m, H4′), 5.87 (1H, d, J = 7 Hz, H1′), 7.56 (1H, s, H6), 7.54–7.64 (4H, m, Ph), 11.54 (1H, s, NH). UV λ_{max} 265, 220; λ_{min} 245. M/z 798.201 $(M+Na)^+$, 814.482 $(M+K)^+$.

1-(2′,3′,5′-Tri-(tert-butyldimethylsilyl)-β-D-ribofuranosyl)-4-triazolo-5-(2-phthalimidooxyethyl)-1H-pyrimidin-2-one (9). To a solution of 1,2,4-triazole (4.8 g, 6.95 mmol) in dry acetonitrile (75 ml) at 0°C was added phosphorus oxychloride (1.3 ml, 1.4 mmol) and the solution stirred at 0°C for 15 min. To this was then added triethylamine (11.6 ml, 8.3 mmol) and the solution stirred for a further 15 min at 0°C. To the solution was then added a solution of 2′,3′,5′-tri-(tert-butyldimethylsilyl)-5-(2-phthalimidooxyethyl)-uridine (**8**) (3.6 g, 4.6 mmol) in acetonitrile (25 ml) and the solution stirred at room temperature overnight (product has same R_f as starting material). The solution was evaporated and worked up as described and then chromatographed ($CHCl_3$/1% MeOH) to give an off-white foam. Yield 2.44 g, 64%. ^{1}H NMR δ (p.p.m.) –0.11–0.12 (18H, m, 6× $SiCH_3$), 0.83–0.91 (27H, m, 3× $C(CH_3)_3$), 3.23–3.30 (2H, m, CH_2CH_2ON), 3.79–3.83 (2H, m, CH_2CH_2ON), 3.93–4.02 (1H, m, H2′), 4.03–4.08 (1H, br s, H3′), 4.21–4.38 (2H, m, H5′, H5″), 4.38–4.40 (1H, m, H4′), 5.87 (1H, d, J = 4 Hz, H1′), 7.51–7.64 (5H, m, H6, Ph), 8.19 (1H, s, triazole CH), 9.36 (1H, s, triazole CH). UV λ_{max} (nm) (10%

MeOH/H_2O) 331, 264. pH 12 λ_{max} 267. M/z 828.696 $(M+H)^+$, 849.629 $(M+Na)^+$, 865.969 $(M+K)^+$.

6-(2′3′5′-Tri-(tert-butyldimethylsilyl)-β-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido[4,5-c][1,2]oxazin-7-one (10). The above triazole (**9**) (2.0 g, 2.4 mmol) was dissolved in dioxane saturated ammonia (25 ml) and the solution stirred at room temperature overnight. The solution was evaporated and the product chromatographed ($CHCl_3$/2% MeOH) to give an off-white foam. Yield 0.88 g, 58%. ^1H NMR δ (p.p.m.) –0.06–0.09 (18H, m, 6× $SiCH_3$), 0.81–0.89 (27H, m, 3× $C(CH_3)_3$), 3.31–3.34 (2H, m, CH_2CH_2ON), 3.66–3.74 (2H, m, H5′, H5″), 3.77–3.87 (3H, m, H2′, CH_2CH_2ON), 3.99–4.01 (1H, m, H3′), 4.09–4.13 (1H, m, H4′), 5.83 (1H, d, J = 7.5 Hz, H1′), 6.79 (1H, s, H6), 10.63 (1H, s, NH). UV λ_{max} (nm) (MeOH) 298 (ε = 7400); λ_{min} (nm) 262. pH 1 λ_{max} 304 (ε = 12 400). pH 12 λ_{max} 303 (ε = 7700). M/z 651.697 $(M + Na)^+$.

6-(β-D-Ribofuranosyl)-3,4-dihydro-8H-pyrimido[4,5-c][1,2]oxazin-7-one (3). The above product (**10**, 0.85 g, 1.35 mmol) was dissolved in methanol (25 ml) and ammonium fluoride (0.3 g, 8.1 mmol) added and then the solution heated at 50°C overnight. The solvent was removed and the product chromatographed ($CHCl_3$/20% MeOH) to give a white solid. Yield 0.31 g, 80%. ^1H NMR δ (p.p.m.) 3.15 (2H, d, J = 5 Hz, CH_2CH_2ON), 3.45–3.58 (2H, m, H5′, H5″), 3.74–3.76 (1H, m, H2′), 3.82 (2H, t, J = 5 Hz, CH_2CH_2ON), 3.91–3.97 (2H, m, H3′,H4′), 4.98–5.01 (2H, m, 2× OH), 5.22 (1H, d, OH), 5.72 (1H, d, J = 5.9 Hz, H1′), 7.00 (1H, s, H6), 10.50 (1H, s, NH). UV λ_{max} (nm) (H_2O) 295 (ε = 5100), 231 (ε = 6100), λ_{min} 260. pH 1 λ_{max} 301 (ε = 7800). pH 12 λ_{max} 302 (ε = 5000). ε260 (μM) 2.5. [cf dP: λ_{max} 295 (ε = 6300), 231 (ε = 7000). pH 1 λ_{max} 301 (ε = 10 000), 224 (ε = 4555). pH 12 λ_{max} 302 (ε = 6400)]. m/z 286.727 $(M+H)^+$, 308.711 $(M+Na)^+$.

6-(β-D-Ribofuranosyl)-3,4-dihydro-8H-pyrimido[4,5-c][1,2]oxazin-7-one-5′-triphosphate (rPTP). To a solution of the nucleoside **3** (80 mg, 0.28 mmol) in trimethyl phosphate (0.5 ml) and triethyl phosphate (0.5 ml) under nitrogen at 0°C was added phosphoryl chloride (36.7 μl, 0.4 mmol), the reaction stirred at 0°C for 3 h. To the solution was then added tributylammonium pyrophosphate (3 ml, 0.5 M in DMF) and tributylamine (0.4 ml) and the solution stirred at 0°C for 45 min when the reaction was complete. The reaction was quenched with TEAB (20 ml, 1 M, pH 8.5) and then evaporated to dryness. The product was dissolved in 15 ml water and purified by anion exchange HPLC (buffer A, water; buffer B 0.3 M TEAB pH 8.5, gradient 0–100% B over 80 min), and then further purified by reverse phase HPLC (RP C-18, buffer A, 0.1 M TEAB; buffer B, 0.1 M TEAB, 50% acetonitrile, gradient 0–100% B over 40 min). Yield 75.3 μmol (28%) as 25 mM solution in water. ^{31}P NMR δ (p.p.m.) (D_2O/EDTA) δ –10.71 (d, γ-P), –11.54 (d, α-P), –23.22 (t, β-P). M/z 526.429 $(M+H)^+$.

Melting experiments

Melting transitions were measured at 260 nm in 100 mM sodium phosphate (pH 7) at an oligomer strand concentration of ~1.8 μM. Absorbance versus temperature for each duplex was obtained at a heating and cooling rate of 0.5°C/min, and the melting transitions (T_m) determined as the maxima of the first differential curves with an error of ±1°C. Thermodynamic calculations were carried out as described by Gralla and Crothers (14).

Circular dichroism (CD) measurements

CD measurements were carried out on a Jobin-Yvon Dichrograph CD6 spectrometer. Data was collected at 0.25 nm intervals, and measurements were carried out between 190 and 330 nm at 20°C. Five such runs were averaged, calculated net of buffer and factor 3 smoothed. Samples were prepared with an oligonucleotide concentration of A_{260} = 0.5 in 10 mM sodium phosphate (pH 7) buffer (15), with a path length of 1 mm.

Polymerase reactions

Polymerase incorporation assays were carried out using the Riboprobe® System (Promega) using SP6, T3 and T7 RNA polymerases, and using pGEM® Express Positive Control Template (Promega). Reactions were carried out according to the manufacturers instructions. rPTP was used in place of either CTP or UTP, and at both 1× and 10× NTP concentrations. Products were electrophoretically separated on 1% agarose gels containing ethidium bromide and visualised under UV light.

TAR RNA synthesis

Wild-type TAR and the TAR mutant G_{26}:C_{39} to C:G RNAs were transcribed from the plasmids BT0 and BT76, respectively (16). In these plasmids, the TAR sequence abuts directly the T3 promoter. The plasmids were digested with *Eco*RI prior to the transcription reactions. Transcripts of 60 nucleotides were generated using either the Riboprobe® System (Promega), or for large scale synthesis RiboMAX® system (Promega). Small scale reaction mixtures for TAR RNA synthesis (typically containing 40 mM Tris–HCl, pH 7.9, 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 2 mM DTT, 40 U RNasin, 400 μM NTPs, 1 μg template DNA, 40 U T3 RNA polymerase, and 20 μCi [α-^{32}P]GTP in 50 μl solution) were incubated for 1 h at 37°C. When rPTP was used instead of UTP and/or CTP, the concentration of rPTP in the reaction was 400 μM (1× concentration) or 4 mM. The products were electrophoretically separated using 12% polyacrylamide gels (165 × 200 × 1 mm) containing 7 M urea at 35 W for 1 h.

Large scale reaction mixtures contained 80 mM HEPES–KOH, pH 7.5, 24 mM $MgCl_2$, 2 mM spermidine, 40 mM DTT, 7.5 mM NTPs, 5 μg template DNA, and 10 μl of T3 enzyme mix in 100 μl solution. The synthesis of TAR was also carried out using 10 mM rPTP instead of UTP. The reaction mixtures were incubated at 37°C for 4 h. The products were electrophoretically separated using 6% polyacrylamide gels (350 × 200 × 1 mm) containing 7 M urea at 35 W for 2.5 h. The transcripts were visualised by autoradiography or UV shadowing for small and large scale syntheses respectively. The correct band was cut out and the RNA was extracted with 0.5 M ammonium acetate/1 mM EDTA. The extracts were then desalted using NAP-10 columns (Pharmacia).

Digestion of TAR RNAs

Digests were carried out as described (17), and the nucleosides separated by HPLC on a Waters μBondapak™ RP-C18 (3.9 × 300 mm) column.

RNA bandshifts

RNA bandshifts were carried out according to the methods described previously (16). In summary, 10 000 c.p.m. (~2 nM) of

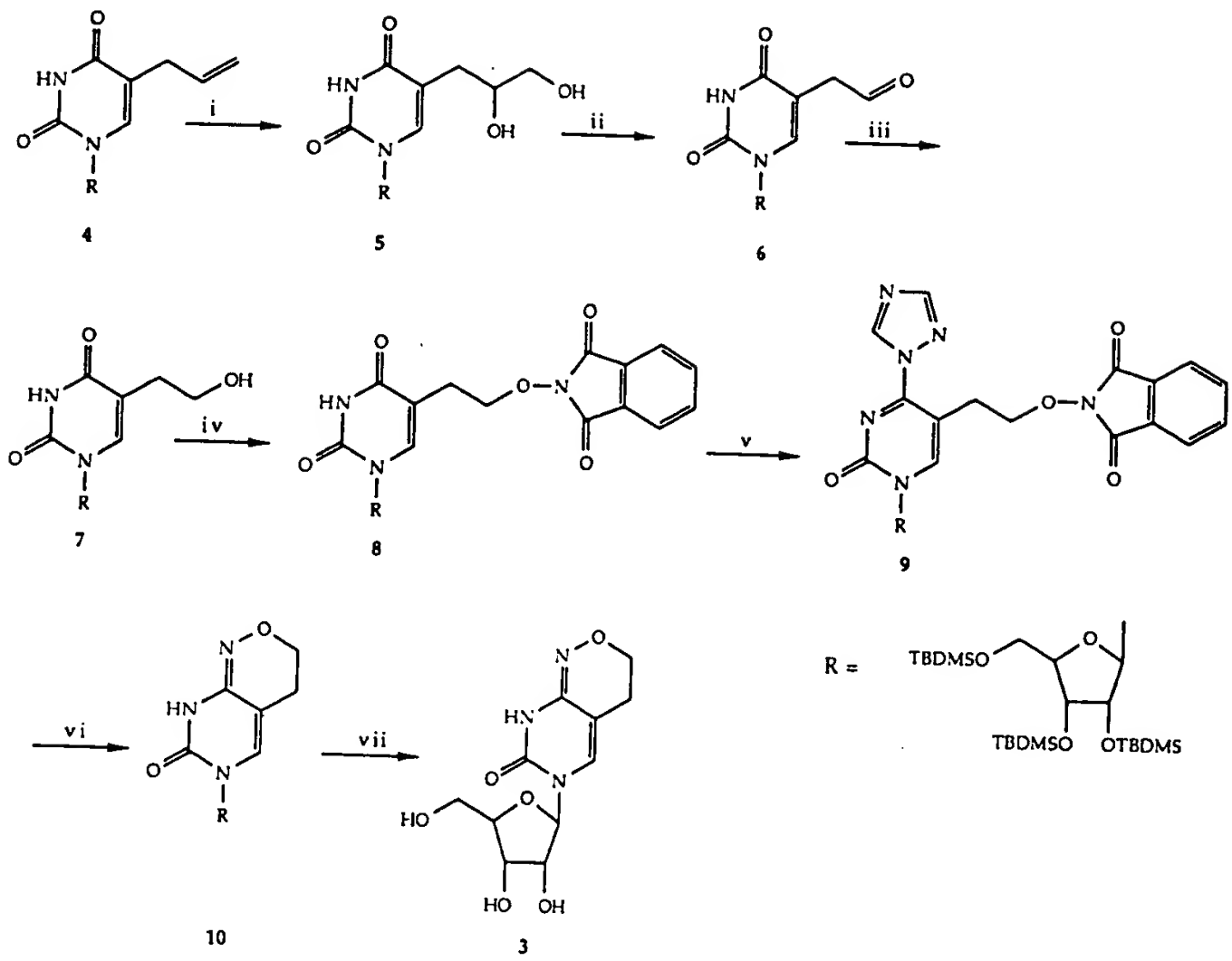


Figure 2. Synthesis of P: (i) *N*-methylmorpholine-*N*-oxide/K_2OsO_4; (ii) $NaIO_4$; (iii) $NaBH_4$; (iv) PPh_3/*N*-hydroxyphthalimide/DIAD; (v) 1,2,4-triazole/$POCl_3$/TEA; (vi) dioxan/NH_3; (vii) NH_4F.

[α^{32}P]GTP labelled TAR RNA (prepared as described above, but using 5-fold excess of rPTP, i.e. 2 mM) was incubated at room temperature for 5 min with various concentrations of ADP-1 peptide (50–1000 nM) in TK buffer (50 mM Tris, pH 7.5, 20 mM KCl) containing 0.1% Triton X-100 and 100 mM DTT. The mixtures were electrophoretically separated (10 W, 1 h) on 8% native polyacrylamide gels (165 × 200 × 1 mm) containing 0.1% Triton X-100. The bands were visualised by overnight exposure to autoradiography film. The autoradiograph was scanned using a Molecular Dynamics Scanning Imager 300A-T.

RESULTS

The synthesis of the ribosyl analogue, rP, is shown in Figure 2. 5-Allyluridine (12), was silylated with *t*-butyldimethylsilyl chloride to give (4) (13). This was then converted to the 5-hydroxyethyluridine (7) by dihydroxylation of the olefinic bond with potassium osmate/*N*-methylmorpholine-*N*-oxide followed by periodate cleavage to the aldehyde (6). Curiously, the periodate reaction was not effective using THF/water or acetone/water, but reaction was obtained using dioxane/water as solvent. The aldehyde 6 was not isolated but converted immediately to the alcohol, 7, by reduction with sodium borohydride. This was then converted to the bicyclic P analogue in a manner similar to that described for the deoxynucleoside (8) by a Mitsunobu reaction of the alcohol (7) with *N*-hydroxyphthalimide, triazolylation and then ring closure using ammonia in dioxane. Finally, the silylated derivative (10) was deprotected using ammonium fluoride in methanol to give the free nucleoside. This was then converted to its 5′-triphosphate derivative.

To screen for the incorporation of the analogue into RNA by the RNA polymerases of the bacteriophages SP6, T3 and T7 the positive control template from the pGEM Express Positive kit (Promega) was used. UTP or CTP were replaced by rPTP in the polymerase transcription reactions. The 2′-deoxynucleoside triphosphate, dPTP, is incorporated opposite dA or dG by *Taq* polymerase in PCR reactions. However, neither TTP or dCTP could be entirely replaced by dPTP (10). Both as a substrate triphosphate and as a template for *Taq* polymerase, dP resembled T more than dC (18). It had also been demonstrated that, in terms of hybridisation, P:A base pairs are equivalent to T:A base pairs, whilst P:G pairs are slightly destabilising when compared with C:G pairs in deoxyribo-oligomers (8). Using T7 polymerase, full length products were formed: when used to replace CTP a distinct, but different, product was obtained. Product yields were lower than the controls using the four NTPs (Fig. 3), but when the rPTP concentration was increased 10-fold compared with the other triphosphates the yield of products was substantially improved. In this instance T3 RNA polymerase incorporated rPTP better in place of CTP rather than in place of UTP (but see below P-TAR transcripts, Fig. 4 and refs 1,3). T7 and SP6 showed a marked preference for replacing UTP rather than CTP with rPTP.

Regulation of HIV-1 transcription is controlled by a specialised RNA/protein interaction (for review see 19). The *trans*-activation responsive region (TAR) of HIV-1 is located immediately downstream of the HIV-1 transcription start site from positions +1

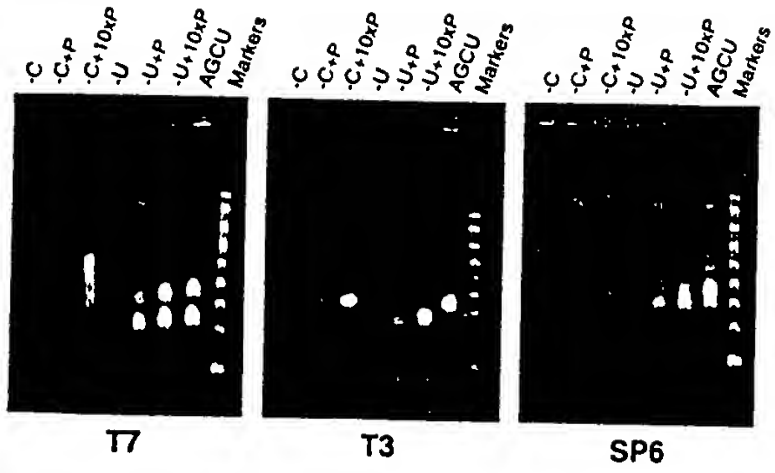


Figure 3. Agarose gel electrophoresis of the products of RNA transcription reactions using the positive control template in the pGEM Express kit (Promega). The products obtained using three separate bacteriophage RNA polymerases T7, T3 and SP6 are shown. Above each lane, the omission of either CTP or UTP (-C or -U) or the addition of rPTP at either the same or 10× the concentration of the other NTPs (+ or +10xP) is recorded.

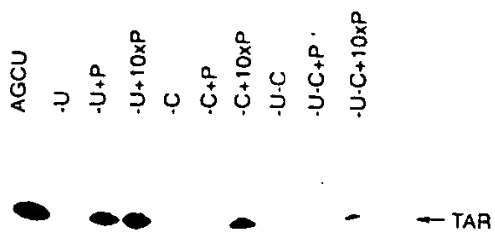


Figure 4. TAR transcripts from wild-type DNA using rPTP to replace UTP (left), CTP or both pyrimidine triphosphates. The figure shows the gel after exposure to film for 30 min and after 2 h where it can be seen that rPTP can be used to entirely replace the pyrimidine triphosphates.

to +59, and is therefore transcribed immediately upon initiation of transcription from the HIV promoter (20). TAR is known to form a highly stable stem–loop and it has a tripyrimidine bulge near the apex of the structure (Fig. 5) (21). TAR RNA is bound by the virally-specified *trans*-activator protein Tat (22). This leads to a dramatic increase in HIV gene expression. The Tat/TAR interaction is very specific and mutations in and around the tripyrimidine bulge abolish both Tat binding and *trans*-activation of HIV transcription (16,23). The binding of Tat to TAR is decreased by the mutation G_{26}:C_{39} to C:G (mGC). A shorter Tat-derived peptide (ADP-1), which contains the core and basic regions of Tat (residues 37–72), has been shown to exhibit similar specificity to the wild-type protein (16,23).

Both wild-type and mGC TAR transcripts were made using the bacteriophage T3 polymerase to assess the efficiency of incorporation of rPTP in place of either UTP or CTP. When transcription was carried out using rPTP to replace UTP, full length RNA transcripts were readily obtained. When replacing

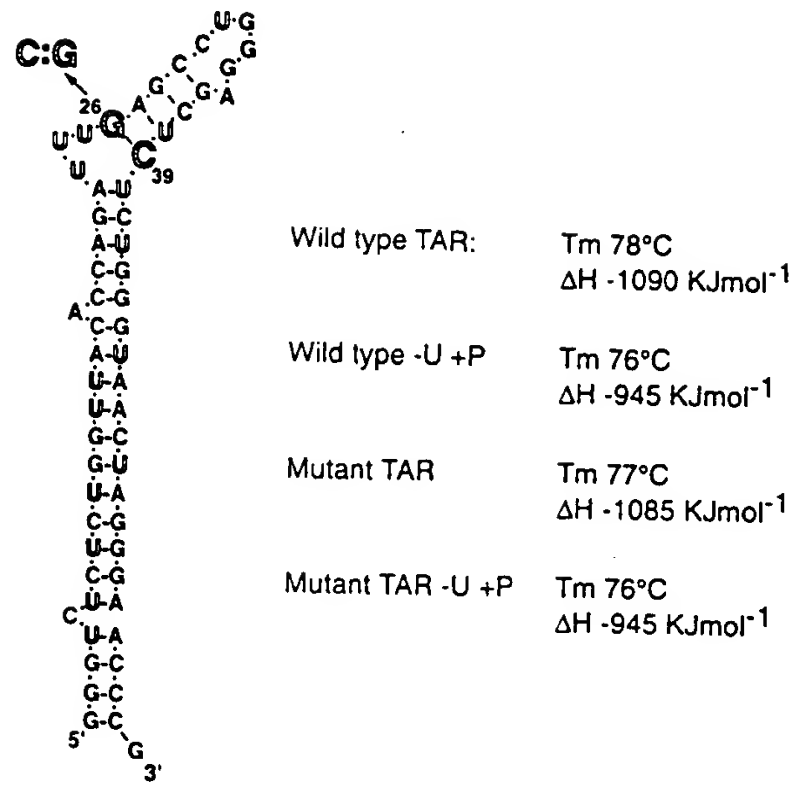


Figure 5. Melting temperatures of each of the TAR RNA transcripts (wild-type and the mGC mutant) synthesised in the absence and presence of rPTP in place of UTP. The figure shows the structure of the wild-type TAR RNA with each of the uridine residues outlined showing the positions where it is known that P is incorporated.

CTP, the yield of transcript was markedly reduced (Fig. 4), but full length transcripts could still be visualised on gels even when both CTP and UTP were replaced by rPTP. Although the reason for this low efficiency of incorporation into a smaller transcription product is unclear, the fact that TAR RNA is a highly ordered structure may provide a possible explanation.

In order to characterise the transcripts, large scale reactions were carried out. TAR RNA itself has been shown to have a high melting temperature, 65°C in 10 mM potassium phosphate, 50 mM sodium chloride (24) consequent on its hairpin structure, and this correlated with our findings (T_m in 100 mM sodium phosphate, 78°C). The transcripts containing rPTP in place of UTP also had a high melting temperature (76°C). This was unexpected in that P:G base-pairs in DNA duplexes are ~2°C less stable per modification than a C:G base-pair (8); there are also two U:G base-pairs in the native TAR RNA which will be replaced by P:G base-pairs in the analogous P-containing TAR (hereafter called P-TAR). However, the derived stacking enthalpy was lower than that of the native TAR; these results are shown in Figure 5. Similar findings were obtained with the mutant (G_{26}:C_{39} to C:G) TAR transcription products. The two P-TAR transcripts also have the same melting temperature and very similar stacking enthalpies. The TAR transcripts were also digested with snake venom phosphodiesterase and alkaline phosphatase to determine their nucleoside composition. The P-TAR transcript was shown to contain only C, G, A and P, though the exact composition could not be accurately determined (data not shown), and therefore it was not possible to show unequivocally that rPTP had also been incorporated in place of CTP, although it might be expected that it would be, albeit at a low frequency and randomly.

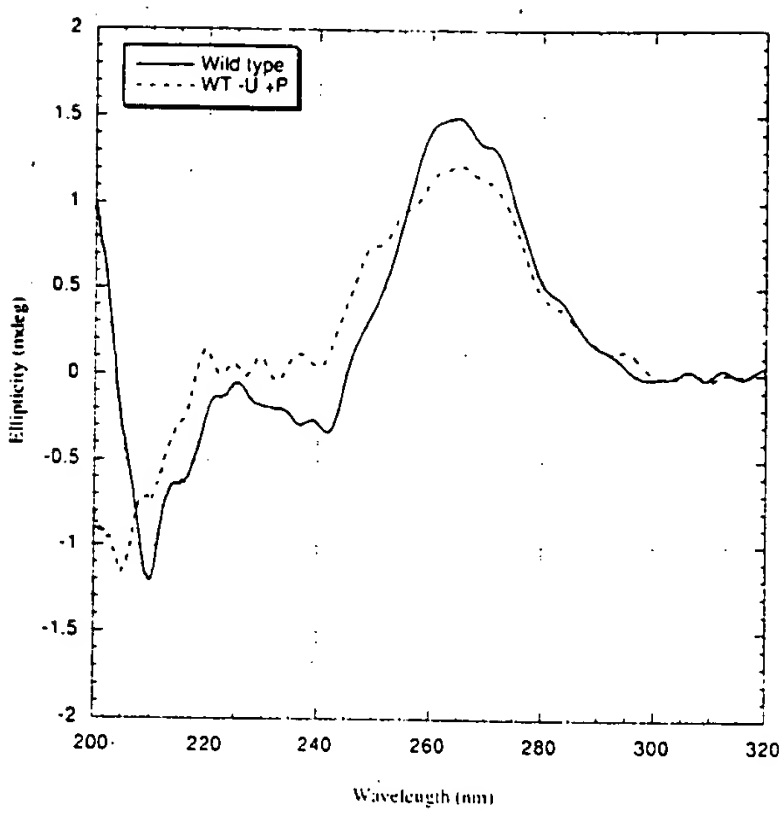


Figure 6. CD spectra of wild type TAR RNA and the P-TAR RNA.

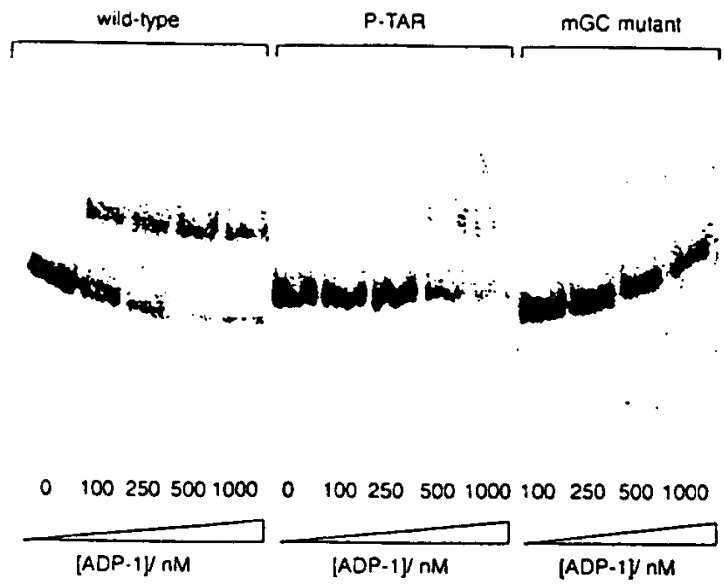


Figure 7. Bandshift of TAR RNA transcripts with ADP-1 peptide at 0–1000 nM ADP-1 concentration. Left shows the bandshift for the wild-type TAR, centre for the P-TAR and right the mGC mutant.

Circular dichroism measurements were carried out on the transcripts. The native TAR RNA had a CD spectrum similar to that previously described (15,24); the P-TAR produced a rather similar spectrum (Fig. 6), although there was a decrease in intensity of the 265 nm band accompanied by a red shift in the crossover and a decrease in intensity of the far UV region bands present in the native structure. Spectra for the mGC mutant TAR and its P-derivative were similar (data not shown).

Binding of the Tat fragment, ADP-1, a fragment of the tat protein (residues 37–72), to the modified TAR transcript was studied. Wild-type TAR RNA, labelled with [α-^{32}P]GTP was synthesised using either UTP or rPTP. The mutant mGC TAR was prepared as a control. RNA bandshift assays were performed in the presence of increasing concentrations of the tat fragment ADP-1 (0–1000 nM). The modified P-TAR ran with a slightly lower mobility due to the PMP residues; this altered mobility has also been observed with oligodeoxyribonucleotides containing dP. The mutant mGC TAR is known to have a 9-fold lower affinity for ADP-1 than the wild-type (16). Figure 7 shows that the affinity of P-TAR for ADP-1 is lower than that of the wild-type TAR but, remarkably, is higher than that of the mGC TAR mutant, and this was confirmed using densitometry measurements. The artefactual second shifted band, known to occur at high concentrations of ADP-1 peptide (16), is also seen with the P-TAR transcript.

DISCUSSION

The ribo-P-5′-triphosphate analogue was incorporated into RNA by all three RNA polymerases T3 and T7, and to a lesser extent for SP6. In the control system (pGEM Express Positive control template) rPTP was successfully used to replace entirely one of the two natural pyrimidine triphosphates to obtain full length transcripts (1.5 kb). In the TAR system using T3 polymerase it proved to be more difficult to obtain a high yield of product when using rPTP in place of CTP. Nevertheless, it was possible to obtain transcripts when both pyrimidine triphosphates were replaced by rPTP if the rPTP concentration was raised 10-fold. When rPTP was used to replace UTP there was a decrease in net synthesis between this and RNA derived by using the four natural triphosphates, but the yield was acceptable.

The replacement of UTP in TAR RNA by rPTP affects a significant number of base-pairs. There are eight U:A base-pairs in addition to the three uridine residues involved in the bulge and the single uridine in the hairpin. It has been shown that of these U23, the first base in the bulge, is essential for recognition by either Tat protein or the peptide ADP-1 (25). C^5-Substituents are tolerated, and therefore it might be expected that the analogue P could be accommodated in this position. Remarkably, binding in the band-shift assay was only reduced by a factor of four compared with the wild-type TAR. In contrast, the biologically inactive mutant (mGC) showed 15-fold reduction in binding (Fig. 7), consistent with the earlier finding that mGC binding was reduced 9-fold (16). It is therefore clear that the P-TAR does bind to ADP-1, demonstrating that P can be tolerated not only throughout the TAR structure, but more importantly in replacing the essential uridine at position 23. The observations provide a *prima facie* case for assuming that the proposed U23:arginine interaction in the TAR:ADP-1 complex (26) can be mimicked by a P23:arginine contact. Overall the experiments indicate that P-TAR has a very similar structure to the wild-type TAR transcript and strongly suggests that it can undergo the large conformational change that occurs on binding (27).

In addition to the nine U:A base-pairs there are also two U:G base-pairs in the native TAR RNA stem. In DNA duplexes the P:G base-pair is in rapid chemical exchange between Watson–Crick and wobble configurations with a very low free energy difference between them (28,29). Evidently the U:G base-pairs in TAR are correspondingly replaced by P:G with very little effect on structure.

The similarity between the wild-type structure and P-TAR was further demonstrated by the fact that they have similar melting temperatures, and CD spectra. They also showed virtual identity in T_m, stacking enthalpy and CD spectra.

Earlier work on the 2′-deoxynucleotide analogue of P (dPTP) demonstrated that it had very similar kinetic parameters of incorporation to those of TTP (10). Further work is in progress to investigate the kinetics of incorporation of the ribo-P triphosphate, and the hybridisation properties of rP-containing oligomers as well as their templating properties.

Work during recent years has shown that interactions in RNA secondary and tertiary structures are much richer than those in DNA and that correspondingly RNA–protein interactions are likewise complex and only now beginning to be understood. We have used the Tat (ADP-1)/TAR system purely as a model to examine the incorporation of rPTP into a biologically active RNA oligomer, without necessarily expecting to learn anything new about the Tat/TAR interaction. Nevertheless, our findings have shown that incorporation of rPTP into TAR has had very little effect on its physical characteristics and biological activity. We are therefore in the process of examining the effect of site specific incorporation of P by chemical synthesis into TAR to examine its effects more rigorously. We believe that these aspects of RNA chemistry can be modulated by the introduction of hydrogen bond degenerate base residues which, in turn, should lead to a variety of novel outcomes and applications.

ACKNOWLEDGEMENTS

We thank Richard Grenfell and Jan Fogg for oligonucleotide synthesis, Dr Rita Bazzanini for mass spectra, and Amersham International plc (to DL), the Isaac Newton Trust (to FH) and the British Council and the Medical Research Council (to KM) for financial assistance.

REFERENCES

1 Banks,G.R., Brown,D.M., Streeter,D.G. and Grossman,L. (1971) *J. Mol. Biol.*, **60**, 425–439.
2 Budowsky,E.I., Sverdlov,E.D. and Spasokukotskaya,T.N. (1972) *Biochim. Biophys. Acta*, **287**, 195–210.
3 Flavell,R.A., Sabo,D.L., Bandle,E.F. and Weissmann,C. (1974) *J. Mol. Biol.*, **89**, 255–272.
4 Müller,W., Weber,H., Meyer,F. and Weissmann,C. (1978) *J. Mol. Biol.*, **124**, 343–358.
5 Brown,D.M., Hewlins,M.J.E. and Schell,P. (1968) *J. Chem. Soc.* (C), 1925–1929.
6 Morozov,Y.U., Savin,F.A., Chekhov,V.O., Budowsky,E.I. and Yakovlev,D.Y. (1982) *J. Photochem.*, **20**, 229–252.
7 Loakes,D. and Brown,D.M. (1994) *Nucleosides Nucleotides*, **13**, 679–688.
8 Kong Thoo Lin,P. and Brown,D.M. (1989) *Nucleic Acids Res.*, **17**, 10373–10383.
9 Kong Thoo Lin,P. and Brown,D.M. (1992) *Nucleic Acids Res.*, **20**, 5149–5152.
10 Zaccolo,M., Williams,D.M., Brown,D.M. and Gherardi,E. (1996) *J. Mol. Biol.*, **255**, 589–603.
11 Hill,F., Loakes,D. and Brown,D.M. (1997) *Nucleosides Nucleotides*, **16**, 1507–1511.
12 Ruth,J.L. and Bergstrom,D.E. (1978) *J. Org. Chem.*, **43**, 2870–2876.
13 Hayakawa,H., Tanaka,H., Obi,K., Itoh,M. and Miyasaka,T. (1987) *Tetrahedron Lett.*, **28**, 87–90.
14 Gralla,J. and Crothers,D.M. (1973) *J. Mol. Biol.*, **78**, 301–319.
15 Loret,E.P., Georgel,P., Johnson,W.C. and Ho,P.S. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 9734–9738.
16 Churcher,M.J., Lamont,C., Hamy,F., Dingwall,C., Green,S.M., Lowe,A.D., Butler,P.J.G., Gait,M.J. and Karn,J. (1993) *J. Mol. Biol.*, **230**, 90–110.
17 Eadie,J.S., McBride,L.J., Efcavitch,J.W., Hoff,L.B. and Cathcart,R. (1987) *Anal. Biochem.*, **165**, 442–447.
18 Hill,F., Loakes,D. and Brown,D.M. (1998) *Proc. Natl. Acad. Sci. USA*, in press.
19 Jones,K.A. and Peterlin,B.M. (1994) *Annu. Rev. Biochem.*, **63**, 717–743.
20 Muesing,M.A., Smith,D.H. and Capon,D.J. (1987) *Cell*, **48**, 691–701.
21 Hauber,J., Perkins,A., Heimer,E.P. and Cullen,B.R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6364–6368.
22 Dingwall,C., Emberg,I., Gait,M.J., Green,S.M., Heaphy,S., Karn,J., Lowe,A.D., Singh,M., Skinner,M.A. and Valerio,R. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6925–6929.
23 Weeks,K.M., Ampe,C., Schultz,S.C., Steitz,T.A. and Crothers,D.M. (1990) *Science*, **249**, 1281–1285.
24 Metzger,A.U., Schindler,T., Willbold,D., Kraft,M., Steegborn,C., Volkmann,A., Frank,R.W. and Rosch,P. (1996) *FEBS Lett.*, **384**, 255–259.
25 Sumner-Smith,M., Roy,S., Barnett,R., Reid,L.S., Kuperman,R., Delling,U. and Sonenberg,N. (1991) *J. Virol.*, **65**, 5196–5202.
26 Aboul-ela,F., Karn,J. and Varani,G. (1995) *J. Mol. Biol.*, **253**, 313–332.
27 Aboul-ela,F., Karn,J. and Varani,G. (1996) *Nucleic Acids Res.*, **24**, 3974–3981.
28 Nedderman,A.N.R., Stone,M.J., Williams,D.H., Kong Thoo Lin,P. and Brown,D.M. (1993) *J. Mol. Biol.*, **230**, 1068–1076.
29 Moore,M.H., Van Meervelt,L., Salisbury,S.A., Kong Thoo Lin,P. and Brown,D.M. (1995) *J. Mol. Biol.*, **251**, 665–673.

Nucleic Acids Research, 1998, Vol. 26, No. 19 4309–4314

SURVEY AND SUMMARY

The use of diaminopurine to investigate structural properties of nucleic acids and molecular recognition between ligands and DNA

Christian Bailly* and Michael J. Waring[1]

INSERM U-124 et Laboratoire de Pharmacologie Antitumorale du Centre Oscar Lambret, IRCL, Place de Verdun, 59045 Lille, France and [1]Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, UK

Received April 7, 1998; Revised and Accepted June 22, 1998

ABSTRACT

2,6-Diaminopurine (DAP) is an analogue of adenine which can be converted to nucleotides that serve as substrates for incorporation into nucleic acids by polymerases in place of (d)AMP. It pairs with thymidine (or uracil), engaging in three hydrogen bonds of the Watson–Crick type. The result of DAP incorporation is to add considerable stability to the double helix and to impart other structural features, such as an altered groove width and disruption of the normal spine of hydration. DNA containing DAP may or may not be recognized by restriction endonucleases; RNA containing DAP may not engage in normal splicing. The DAP·T pair affects the local flexibility of DNA and impedes the interaction with helix bending proteins. By providing a non-canonical hydrogen bond donor in the minor groove and/or blocking access to the floor of that groove it strongly affects interactions with small molecules such as antibiotics and anticancer drugs. Examples which illustrate altered recognition of nucleotide sequences in DAP-containing DNA are presented: changed sites of cutting by bleomycin, photocleavage by uranyl nitrate and footprinting with mithramycin. Using DNA in which both A→DAP and G→Inosine substitutions have been made it is possible to assess precisely the role of the purine 2-amino group in ligand–DNA recognition.

INTRODUCTION

It was exactly half a century ago when the first modified nucleobase, 5-methylcytosine (5-MeC), was discovered (1). Nowadays, it is well established that this unusual but naturally occurring base participates in the control of gene expression in higher organisms (2). Over the last 50 years many other modified bases have been discovered in DNA. One of the most interesting is 2-aminoadenine (abbreviated to DAP or D for 2,6-diaminopurine), which is used in place of adenine by the cyanophage S-2L (3,4), whose very existence shows that DAP in DNA is compatible with normal DNA function (5,6).

The DAP·T base pair possesses an extra hydrogen bond compared with A·T because of the additional -NH_2 pointing toward the minor groove of DNA (Fig. 1). DAP is a common tool in nucleic acid chemistry which can be used to study molecular recognition between DNA or RNA and ligands, both small and large. In this paper, various applications of DAP are briefly presented. In particular, the emphasis is on how this base can be extremely useful in investigations on the structure of nucleic acids as well as sequence-specific interactions between DNA and small molecules or proteins.

CHEMICAL AND ENZYMATIC SYNTHESIS OF DIAMINOPURINE-CONTAINING DNA

In most cases, the DAP base is introduced chemically into DNA sequences using conventional phosphoramidite chemistry. Synthesis of the DAP nucleoside phosphoramidite has been described (7). Alternatively, DAP can be incorporated into DNA by enzymatic methods via the use of DAP triphosphate and polymerases. The triphosphate of 2-aminoadenosine acts as a true analogue of ATP in transcription (8). 2,6-Diaminopurine-2′-deoxyribonucleoside is commercially available and methods have been described to convert it into dDTP. The synthesis involves converting 2-amino-2′-deoxyadenosine to its 5′-monophosphate derivative dDMP followed by pyrophosphorylation (9,10). Other chemical routes have been reported (11,12). Alternatively, the triphosphate nucleotide dDTP may be produced biotechnologically directly from the base precursor DAP, as is the case with the related nucleobase 2-aminopurine. When added to growing bacteria, 2-aminopurine is metabolized to form deoxy-2-aminopurine triphosphate (dAPTP) (13) which, like dDTP, serves as a substrate for DNA polymerases (14,15). It is known that DAP, which is toxic to cultured cells, is normally metabolized to DAP ribonucleoside and then deaminated to guanosine (16). DAP and its 2′-deoxyriboside (DAPdR) exert their toxicity by different mechanisms. DAPdR, but not DAP, acts as a precursor of deoxyguanosine in mammalian cells (17). The

*To whom correspondence should be addressed. Tel: +33 320 16 92 20; Fax: +33 320 16 92 29; Email: bailly@lille.inserm.fr

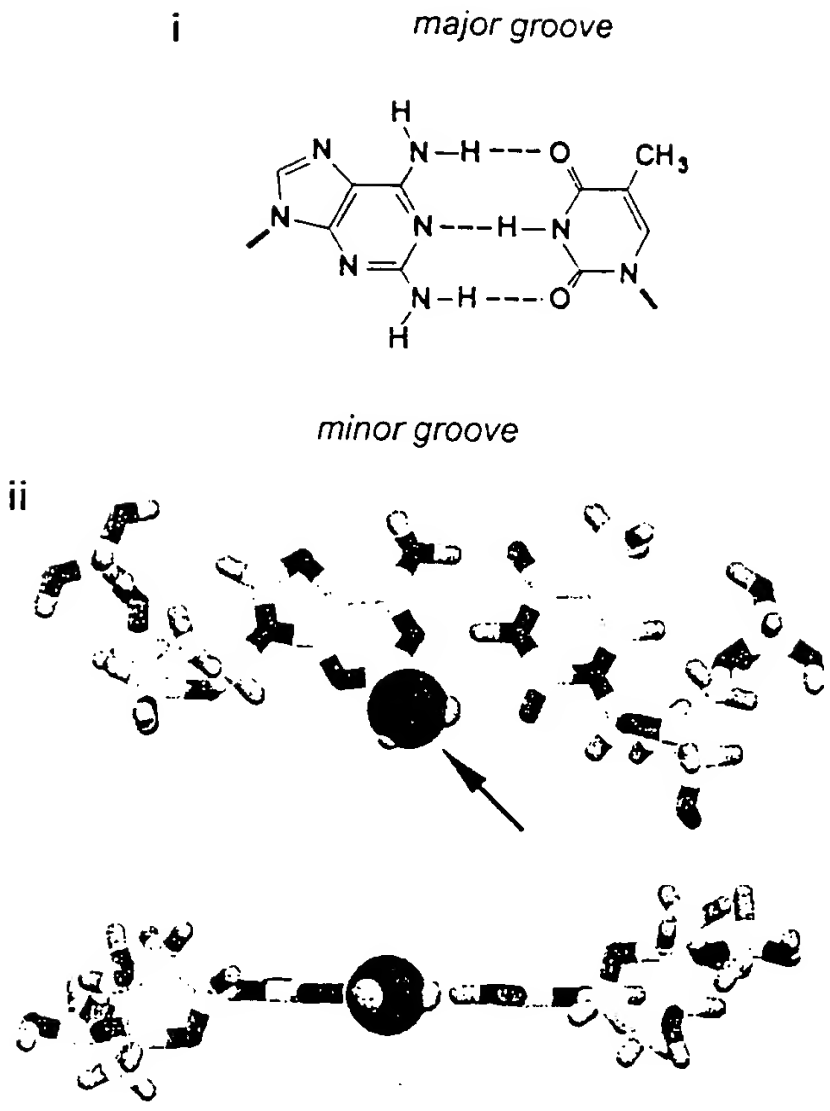


Figure 1. Structure of a diaminopurine-thymidine base pair (Watson–Crick pairing). The positions of the minor and major grooves are indicated. The 2-amino group which distinguishes a D·T pair from an A·T pair is printed in red. Broken lines represent hydrogen bonds. The molecular model was built with the programs HyperChem 5.1 and Alchemy 2000. The guanine 2-amino group is shown as a blue sphere.

polymerization of chemically synthesized DTP by deoxynucleotidyl transferase from calf thymus has been reported (18).

DIAMINOPURINE INCREASES THE THERMAL STABILITIES OF DUPLEX DNA AND RNA

DAP is frequently introduced into nucleic acid sequences to increase the melting temperature of DNA (19) and/or RNA duplexes (20) and for related applications such as primers for sequencing and PCR (21,22) and fingerprinting (23). Incorporation of DAP into short DNA oligomers increases the thermal stability of the duplex by 0–2°C per D·T base pair (24). Prosnyak *et al.* (23) reported that the melting temperature (T_m) of a given oligo or polynucleotide containing y% of DAP is increased by a factor of $0.14y$ (ΔT_m). However, dependence of the T_m on the DAP content is not linear (25,26). For a 160 bp DNA fragment having all A replaced with DAP residues on both strands of the fragment, we measured T_m values of 65.9, 77.4 and 78.6°C, whereas the T_m values for the corresponding DNA fragment containing natural bases were 62.7, 70.0 and 72.8°C (27). The T_m elevation resulting from introduction of a 2-amino group onto A residues is much smaller in the deoxy series than in the ribo series (5). In the deoxyribo series the stabilizing contribution arising from the formation of a third hydrogen bond in D·T pairs is opposed by a destabilization due to disruption of the spine of hydration in the minor groove of B-form DNA (28).

The potential uses of DAP are wide ranging. For example, it has recently been used in conjunction with 2-thiothymine to produce selectively binding complementary (SBC) oligonucleotides which can facilitate the formation of particular DNA structure (e.g. three-arm junctions) via strand invasion (29). DAP, like 2,4-diaminopyrimidine, can be used for the development of components of an extended genetic alphabet (30). Very recently, Nielsen and co-workers reported that incorporation of DAP nucleobases into peptide nucleic acids (PNA) increased the DNA binding and sequence discrimination of PNA oligomers (31).

USE OF DIAMINOPURINE BY DNA POLYMERASES

Incorporation of nucleoside triphosphate analogues by polymerases is a method of choice to examine miscoding by different DNA polymerases. For example, 8-oxo-dGTP and 8-amino-dGTP were employed in a recent study comparing the efficiency of utilization of these modified bases by HIV type-1 and murine leukemia virus reverse transcriptases with that of mammalian DNA polymerases (32). The HIV-1 reverse transcriptase readily accepts non-canonical bases, as well as nucleoside triphosphates modified on the sugar (33). Lutz *et al.* (30) have shown that the AIDS virus enzyme successfully incorporates the triphosphate form of 2,4-diaminopyrimidine opposite 2′-deoxy-7-deazaxanthosine, whereas the same reaction failed completely with calf thymus DNA polymerases or with the Klenow fragment of *Escherichia coli* DNA polymerase I. As regards DAP, it is a good substrate for a number of polymerases. Substantial changes in the minor groove of DNA do not disrupt recognition contacts with T3 or T7 RNA polymerases. Apparently, there is no interaction between RNA polymerase and the guanine 2-amino group (34,35). Heat stable polymerases can also function with modified base-containing nucleoside triphosphates. dDTP is readily accepted by *Taq* polymerase and related enzymes. We have shown that incorporation of the diaminopurine nucleobase into both strands of a 160mer fragment presents no particular difficulties (36).

STRUCTURAL STUDIES OF DIAMINOPURINE-CONTAINING DNA

Structural studies of DNA have also profited from the use of DAP. Crystal structures of the hexanucleotides d(CGUDCG)$_2$, d(CGTDCG)$_2$ and d(CDCGTG)$_2$ revealed that substitution with a central D·U or D·T pair is consistent with presumed Z-DNA formation (37,38). The net effect of adding an N2 amino group to the C2 carbon of the adenine base in a T·A base pair is to render the minor groove of both B- and Z-DNA more hydrophilic. Although the effect of the added exocyclic amino group on the stability of a Z conformer is greater than for a B conformer, both the Z and B structures have a C2 carbon that becomes almost entirely inaccessible to solvent upon addition of the N2-amino group (39). Structural studies with DAP-containing oligonucleotides (and other modified bases, in particular inosine) have shown unambiguously that the N2-amino group is critically important in defining the stability of Z- versus B-DNA. It is interesting to note that with the Z-form oligonucleotide d(CDCGTG), just as with the hexanucleotide d(CGCGCG), the continuous spine of water molecules in the minor groove crevice is not disrupted, suggesting that these sets of water molecules help to stabilize Z-DNA conformation (37,39).

The effect of the purine 2-amino group on the preferred conformational states of DNA has also been studied with the synthetic polymer poly(dD-dT)$_2$ (28,40). Spectroscopic analysis of poly(dD-dT) showed that a salt-dependent transition from the standard B-form to an unusual A-form conformation occurs under certain solvent conditions and/or in the presence of polyamines (41–43). The putative A-form of poly(dD-dT) is stabilized by the methyl group at position 5 of the pyrimidine base (44). However, comparative NMR analysis of the dodecanucleotide duplex d(GCATTATTGC) and its analogue having all A replaced by D indicated that the A→DAP substitution does not disturb the global or local conformation of the DNA duplex, at least under low salt conditions (45).

Uranyl nitrate is a sensitive tool for probing the structure of nucleic acids. The extent of photocleavage of double-stranded DNA by uranyl ions at acidic pH (~6.0–6.5) exhibits a very strong modulation which is correlated with minor groove width/electronegative potential (46). The width of the minor groove of the DNA helix is to a first approximation correlated with its AT/GC content since AT-tracts appear to be associated with a narrowed groove, whereas GC-tracts have a widened minor groove. Studies of uranyl-mediated cleavage of DNA containing DAP residues have revealed that the variation of the minor groove width with the local nucleotide composition can be attributed to a large extent to the presence of a purine 2-amino group on G·C base pairs.

As shown in Figure 2, the reactivity towards uranyl nitrate at acid pH is modulated in the DAP-containing DNA quite differently from natural DNA, consistent with a marked widening at sites of A→DAP replacement. The interpretation is clear, that A→DAP substitutions markedly affect the minor groove width of DNA. The changes in groove width are equally evident in the patterns of susceptibility of these two DNA species to DNase I cleavage (47).

The exocyclic amino group of G or D plays a significant role in the intrinsic curvature of DNA. Gel electrophoresis studies employing a series of oligonucleotides 5′-AAAAAGCCGC-3′ where A and G residues were systematically substituted with D or I residues at different positions indicated that the curvature induced by an A-tract in DNA molecules is primarily located at the junction with the 3′-end of the A-tract (48,49).

DIAMINOPURINE AND RNA STRUCTURE AND FUNCTION

The nucleotide analogue interference mapping assay is a sensitive method by which to identify and determine the exact contribution of the chemical groups within RNA which are essential for its activity (50). The technique has recently been used to investigate the role of every N2-exocyclic amine of G within a large RNA, the *Tetrahymena* group I intron, using diaminopurine riboside monophosphate (50). This DAP riboside also proved valuable for studying G·U wobble pairs in a variety of RNA molecules (51,52) and for investigating spliceosome assembly. Antisense probes incorporating DAP are efficiently able to select RNP particles which would otherwise be inaccessible (20). This suggests that DAP may improve antisense activity.

DIAMINOPURINE AND PROTEIN–DNA RECOGNITION

The 2-amino group of guanine residues is the only hydrogen bond donor group exposed in the minor groove of DNA. In addition, it impedes access to the floor of the groove and it interferes with

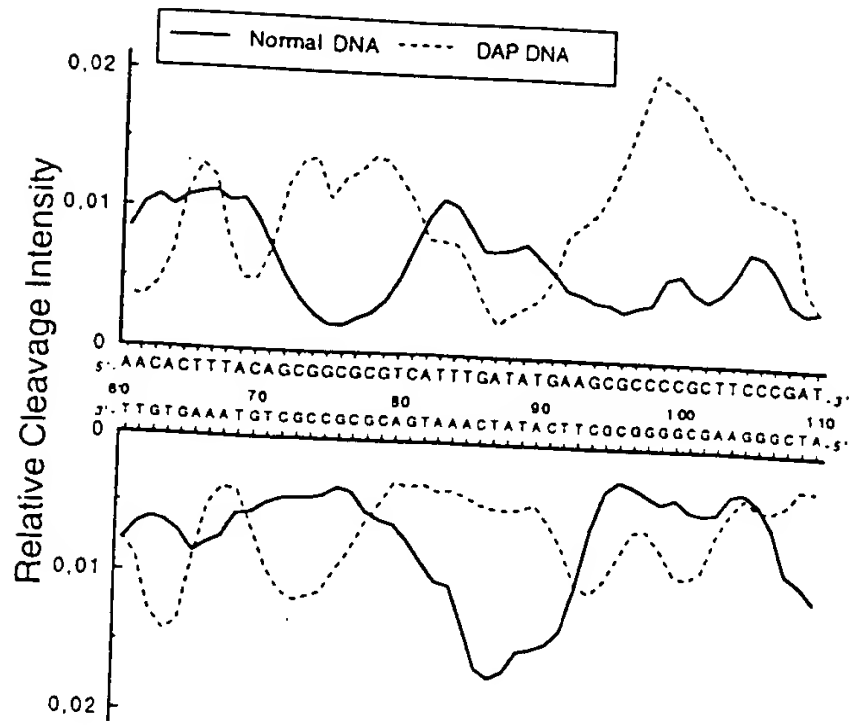


Figure 2. Cleavage plots comparing the susceptibility of the 160 bp *tyrT* DNA fragment containing natural bases or D·T pairs in place of A·T pairs to the uranyl-mediated photocleavage reaction. The double-stranded sequence shown above and below the axis corresponds to that of the normal DNA. In the DAP DNA, all adenine residues are replaced by diaminopurine residues (47).

the spine of hydration. As such, it could be expected to play a determinant role in the recognition of DNA sequences by proteins, peptides and small molecules.

The crystal structure of the restriction endonuclease *Eco*RI complexed to the tridecamer d(TCGCGAATTCGCG) containing the recognition sequence (underlined) reveals that the enzyme makes contact with the double helix mainly in the major groove (53). Specific hydrogen bonding interactions can be detected between the enzyme and the O6/N7 of guanine and N6/N7 of adenine at the target site. There is no apparent specific interaction involving the N2-amino group of guanine in the minor groove of the duplex at the restriction sequence. As a result, one would anticipate little, if any, perturbation of *Eco*RI cleavage of modified DNA duplexes containing bases with altered 2-amino groups. Nevertheless, various experimental studies have shown that replacement of deoxyguanosines or deoxyadenosines in 5′-GAATTC with deoxyinosines or deoxydiaminopurines, respectively, can significantly change the enzymatic activity of *Eco*RI (54,55). For example, substitution with DAP at position 3, d(GADTTC), resulted in a 9-fold decrease in the specificity constant (54). The use of DAP-containing substrates has provided valuable clues to the mechanism by which *Eco*RI and related enzymes (e.g. *Rsr*I; 56) recognize the duplex sequence GAATTC. Similar results have been reported with other restriction endonucleases that mainly bind and cut via the major groove of DNA (10,57). DAP has also been used to study the kinetics of DNA methylation by the *Eco*RI modification methylase and *Dam* methyltransferase (58,59).

Thus, although *Eco*RI can be considered essentially as a major groove-binding protein, the nuclease is evidently also sensitive to the functional groups exposed in the opposite minor groove. The same conclusion was reached when we studied the interaction between the factor for inversion stimulation (FIS) and DNA containing inosine and/or DAP residues. FIS is a major groove-binding protein from *E.coli* required for several processes, including site-specific recombination, transcriptional activation

and DNA replication (60). We have shown that base substitutions which alter the placement and presence of the purine 2-amino group in the minor groove can affect both the intrinsic curvature and the bendability of DNA and thereby modulate the interaction with proteins like FIS, whose binding sites lie purely within the major groove of the double helix (61). Two other proteins whose interaction with DNA is strongly affected by A→D and G→I replacements are HMG-D, a member of the HMG-1 family of chromosomal proteins (62), and the integration host factor (IHF) from *E.coli*, a small protein which binds preferentially to AT-rich sequences. Binding studies with a series of 41 bp oligonucleotides containing adenine analogues revealed that the interaction of IHF with the double helix involves contacts within both the minor and major groove. A→D substitution within the binding region considerably reduces the protein binding affinity (up to 50-fold) (63). For both HMG-D and IHF, the effect is attributable to an indirect influence mediated via helix deformability.

DIAMINOPURINE AND DNA REPAIR

Owing to the structural similarities between D·T and G·T pairs, DAP has been used to study the repair of G·T mismatches. Experiments using cell-free extracts and 45 bp oligonucleotides containing D·T pairs at defined positions suggest that the repair mechanism operating on D·T pairs may be the same as the human G·T repair pathway (64). However, more recent *in vitro* studies have revealed marked differences as regards the extent of incision by human thymine glycosylase of 45 bp heteroduplexes bearing G·T mispairs or D·T pairs (65).

DRUG–DNA RECOGNITION STUDIES USING DIAMINOPURINE

About 30 years ago the idea was first proposed that the 2-amino group of guanine is a significant determinant for the binding of small molecules within the minor groove of the double helix. Cerami *et al.* (66) studied binding of the antitumour drug actinomycin (an antibiotic that remains extensively used in cancer chemotherapy) to poly(dI)·(dC) and to a synthetic analogue of poly(dA-dT)·(dA-dT) containing DAP bases partly or wholly replacing the adenine bases. This was the first experiment using polydeoxynucleotides in which the purine 2-amino group was deleted (G→I substitution) or added to adenine residues (A→D substitution). Later, short oligonucleotides containing DAP residues were used as substrates for DNA binding/cleavage studies. Using hexanucleotides possessing A·T, G·C, I·C or D·T pairs, Sugiyama *et al.* (67) investigated the mechanism of DNA cleavage by the antibiotic neocarzinostatin. Gao *et al.* (68) also used a hexanucleotide, d(CGTDCG)$_2$, to refine the X-ray structure of the covalent formaldehyde-mediated complex between this DAP-containing duplex and the daunorubicin derivative MAR-70.

Since the pioneer work of Cerami *et al.* (66), the influence of the 2-amino group of guanine on drug–DNA recognition has always been considered critical, but its exact role has remained uncertain. It is only recently that its precise function has been elucidated, through the combined use of PCR technology and base-modified nucleoside triphosphates, which enabled us to determine how and to what extent the 2-amino group of guanine contributes to the recognition of specific DNA sequences by a large diversity of small molecules.

Sequence-specific cleavage of DNA by the antitumour antibiotics bleomycin and calicheamicin γ_1^I is strongly dependent on the position of the purine 2-amino group. For bleomycin, relocating the 2-amino group from guanine to adenine nucleotides creates new cleavage sites at pyrimidine residues lying 3′ of DAP residues. For calicheamicin, the presence of a purine 2-amino group adjacent to the cutting site potentiates the cleavage reaction (69). The 2-amino group also constitutes a key structural element for sequence-specific recognition of DNA by non-covalent binders. Irrespective of their mode of interaction with the double helix, such GC-specific antibiotics as mithramycin and chromomycin find new binding sites associated with DAP-containing sequences in (I+DAP)-substituted DNA and are excluded from former canonical sites containing I·C base pairs. The converse was found to be the case for a group of normally AT-selective ligands which bind in the minor groove of the helix, such as netropsin, berenil and DAPI (36,62,70). The binding sites of almost all DNA-binding drugs and antibiotics strictly follow the placement of the purine 2-amino group, which serves as both a positive and negative effector (69).

To illustrate the effect of DAP residues on drug–DNA recognition, a footprinting gel obtained with the antitumour antibiotic mithramycin is presented in Figure 3. The effect of shifting the purine 2-amino group from guanines to adenines (by virtue of combined A→D and G→I substitutions) is to provoke a significant redistribution of binding sites for mithramycin, such that the newly created sites containing D·T pairs are substantially preferred over G·C-containing sites (36). For example, the strong footprint around position 100 with normal DNA in the presence of mithramycin is totally absent with the modified DNA species and, conversely, the footprint around position 87 with I+DAP DNA corresponds to a region of enhanced DNase I cleavage with normal DNA. Also, the strong footprint observed around position 75 with normal DNA is missing in DNA containing both I and DAP residues. There is no doubt that the pattern of binding sites for mithramycin is radically changed. The drug is displaced from its G·C sites in natural DNA to pick up new sites in the DAP·T-rich sequences created in the modified nucleic acid.

The most pronounced effects attributable to DAP were observed with the quinoxaline antibiotics such as echinomycin and triostin A (71,72). The A→D substitution potentiates enormously the interaction of these two drugs with DNA, up to 1000-fold. With normal DNA, a concentration in the 10–20 μM range is needed to evidence strong binding to CpG sites. With the DAP-containing DNA, footprints are already very pronounced at only 0.5 μM drug and binding to newly created TpD sites can be unambiguously detected at a concentration as low as 10–20 nM (72). This enhancement of binding is only seen with the naturally occurring quinoxaline antibiotics and does not occur with the synthetic analogue TANDEM, which recognizes A·T-containing sites and is totally insensitive the relocation of the exocyclic amino group (72). In this case, the binding specificity may arise primarily from stacking and hydrophobic interactions rather than from direct contact with the exocyclic guanine 2-amino group (72). With other drugs, such as actinomycin, the extent of binding to DAP·T sites is essentially unchanged compared with normal DNA (71). There is something special about regions of alternating T·D base pairs which generates unusually good binding sites for echinomycin and triostin A. The local structure and/or the rigidity of the TpD sites could be exploited by the drugs to fit particularly neatly within the minor groove. Alternatively, the stacking of their quinoxaline rings upon DAP·T base pairs could be especially propitious (73).

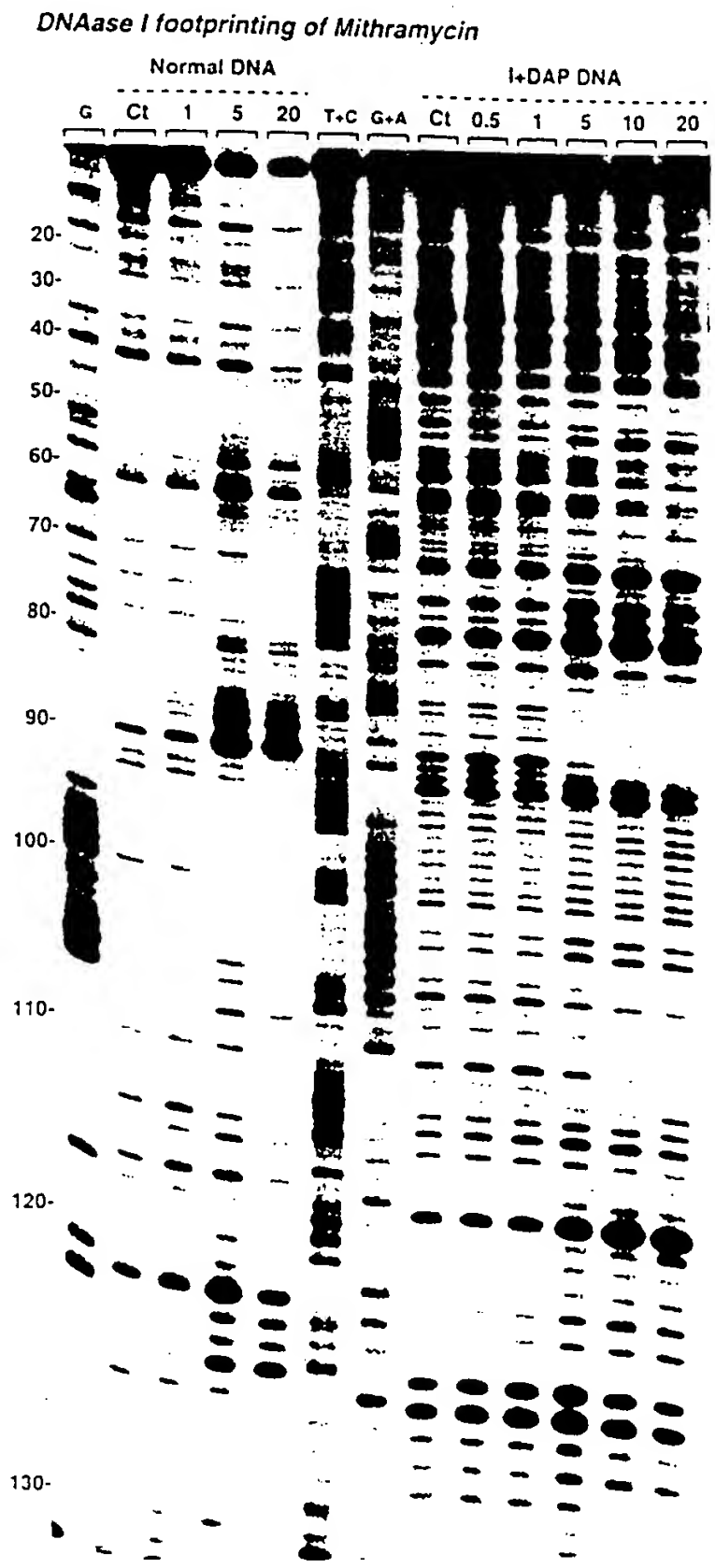


Figure 3. DNase I footprinting of mithramycin on the Crick (sense) strand of normal and inosine + DAP-containing *tyrT*(A93) DNA. The products of DNase I digestion were identified by reference to the Maxam–Gilbert markers (lanes T+C and G+A). Control lanes (Ct) show the products resulting from limited DNase I digestion in the absence of drug. The remaining lanes show the products of digestion in the presence of the indicated antibiotic concentrations (expressed as micromolar). The scale on the left corresponds to the standard numbering of the *tyrT* DNA as represented in Figure 2.

There are very few cases where the replacement of adenine with DAP has little or no effect on the sequence selectivity associated with drug binding. The footprinting profile of the 'threading' intercalator nogalamycin is potentiated in DAP+I-substituted DNA but otherwise remains much the same as with natural DNA (36). Studies with a series of antitumour bis-naphthalimide derivatives which bis-intercalate into DNA sequences, particularly those containing GpT (ApC) and TpG (CpA) steps, showed that repositioning the 2-amino group of G·C base pairs by substitution with inosine and/or DAP had little effect on the distribution of drug molecules between binding sites. In contrast to nearly all common (bis)intercalating drugs, the bis-naphthalimides appear to engage in contacts with the edges of the base pairs via the major groove of the double helix (74).

CONCLUSION

After incorporation of DAP into DNA, the significant difference from adenine is that a 2-amino group is present in the minor groove (Fig. 1). The consequent modification of the surface of the minor groove results in altered conformational properties of the double helix, leading to altered recognition by proteins and small molecules. The D·T base pair is more stable than an A·T pair. The reinforcement of base pairing reduces the flexibility of the DNA and thus generally reduces the extent of protein binding, at least for small proteins such as FIS, HMG-D and IHF. A→D substitutions also decrease the capacity for binding within the minor groove of antibiotics such as netropsin and distamycin, whereas for other drugs, like the quinoxaline antibiotics echinomycin and triostin, the DAP substitution promotes the recognition process considerably. It is likely that the larger D·T pair (having a dipole moment of ~2.3) gives rise to better stacking interactions with an intercalating chromophore than the standard A·T pair (dipole moment ~1.7) (75–77). The changes in surface area and electrostatic properties of the base pair may favour interaction with a planar chromophore but the strength of intercalation must also depend on the interaction of the attached groups (e.g. the peptide moiety of echinomycin or the sugar moiety of calicheamicin) with the sequences flanking the intercalating site. The approach described here, which uses modified bases to study ligand–nucleic acid interactions, provides useful information, but one should bear in mind that the observed effects can arise from direct interaction between the ligand and the newly introduced group on the base, as well as from indirect interactions. Analogue substitution, be it with DAP or another base, can always affect the position or properties of neighbouring bases that might be involved in direct interactions with the ligand.

In addition to the varied applications presented above, DAP has been used for the recognition of abasic sites in DNA (78), to investigate mechanisms of mutagenesis (79,80) and to select purine-resistant variants from mutagenized cultures of *Drosophila* (81). Thus, the potential applications of DAP as a nucleobase cover many modern aspects of nucleic acid chemistry from structure to biology. DAP represents a useful chemical product particularly well suited to increase the stability of double-stranded nucleic acids. At the same time it is a natural product and a key element of the genetic machinery of the S-2L cyanophage. The ability of polymerases to accept non-standard base pairs, such as D·T pairs, is remarkable in the light of the physiological role that the polymerases play. Studies with DAP and other modified bases reinforce the idea that the Watson–Crick formalism can be extended while maintaining a high fidelity of DNA replication which is essential for preserving the integrity of living organisms. DAP-containing nucleic acids can be prepared in many synthetically convenient ways so as to produce different repertoires of molecules having a range of functionalities. There is good reason to believe that the successful use of DAP will continue to inspire the development of novel nucleic acid products and the emergence of new techniques.

ACKNOWLEDGEMENTS

This work was done with the support of research grants (to C.B.) from the Ligue Nationale Française Contre le Cancer (Comité du Nord) and (to M.J.W.) from the Cancer Research Campaign, the European Union and the Association for International Cancer Research. The authors also thank the Sir Halley Stewart Trust. The authors thank Prof. J.B.Chaires (Department of Biochemistry, University of Mississippi, MS) for the drawing of the D·T pair and Prof. F.Gago (Department of Pharmacology, University of Alcala, Alcala, Spain) for useful information on DAP.

REFERENCES

1 Hotchkiss,R.D. (1948) *J. Biol. Chem.*, **168**, 315–322.
2 Cheng,X. (1995) *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 293–318.
3 Kimos,M.D., Khudyakov,I.Y., Alexandrushkina,N.I. and Vanyushin,B.F. (1977) *Nature*, **270**, 369–370.
4 Khudyakov,I.Y., Kimos,M.D., Alexandrushkina,N.I. and Vanyushin,B.F. (1978) *Virology*, **88**, 8–18.
5 Howard,F.B. and Miles,H.T. (1983) *Biochemistry*, **22**, 597–600.
6 Howard,F.B. and Miles,H.T. (1984) *Biochemistry*, **23**, 6723–6732.
7 Sproat,B.S., Iribarren,A.M., Garcia,R.G. and Beijer,B. (1991) *Nucleic Acids Res.*, **19**, 733–738.
8 Rackwitz,H.R. and Scheit,K.H. (1977) *Eur. J. Biochem.*, **72**, 191–200.
9 Kahn,N.N., Wright,G.E., Dudycz,L.W. and Brown,N.C. (1985) *Nucleic Acids Res.*, **13**, 6331–6342.
10 Chollet,A. and Kawashima,E. (1988) *Nucleic Acids Res.*, **16**, 305–317.
11 Brennan,C.A. and Gumport,R.I. (1985) *Nucleic Acids Res.*, **13**, 8665–8684.
12 McLaughlin,L.W., Leong,T., Benseler,F. and Piel,N. (1988) *Nucleic Acids Res.*, **16**, 5631–5644.
13 Rogan,E.G. and Bessman,M.J. (1970) *J. Bacteriol.*, **103**, 622–633.
14 Bessman,M.J., Muzyczka,N., Goodman,M.F. and Schnaar,R.L. (1974) *J. Mol. Biol.*, **88**, 409–421.
15 Muzycka,N., Poland,R.L. and Bessman,M.J. (1972) *J. Biol. Chem.*, **247**, 7116–7122.
16 Garber,B.B. and Gots,J.S. (1980) *J. Bacteriol.*, **143**, 864–871.
17 Weckbecker,G. and Cory,J.G. (1989) *Adv. Enzyme Regulat.*, **28**, 125–144.
18 Scheit,K.H. and Rackwitz,H.R. (1982) *Nucleic Acids Res.*, **10**, 4059–4069.
19 Hoheisel,J.D. and Lehrach,H. (1990) *FEBS Lett.*, **274**, 103–106.
20 Lamm,G.M., Blencowe,B.J., Sproat,B.S., Iribarren,A.M., Ryder,U. and Lamond,A.I. (1991) *Nucleic Acids Res.*, **19**, 3193–3198.
21 Azhykina,T.L., Veselovskaya,S.I., Myasnikov,V.A., Potapov,V.K. and Sverdlov,E.D. (1993) *Proc. Natl Acad. Sci. Russia*, **330**, 624–627.
22 Lebedev,Y., Akopyants,N., Azhikina,T., Shevchenko,Y., Potapov,V., Stecenko,D., Berg,D. and Sverdlov,E. (1996) *Genet. Anal.*, **13**, 15–21.
23 Prosnyak,M.I., Veselovskaya,S.I., Myasnikov,V.A., Efremova,E.J., Potapov,V.K., Limborska,S.A. and Sverdlov,E.D. (1994) *Genomics*, **21**, 490–494.
24 Gryaznov,S. and Schultz,R.G. (1994) *Tetrahedron Lett.*, **35**, 2489–2492.
25 Muraoka,M., Miles,H.T. and Howard,F.B. (1980) *Biochemistry*, **19**, 2429–2439.
26 Sagi,J., Szakonyi,E., Vorlickova,M. and Kypr,J. (1996) *J. Biomol. Struct. Dyn.*, **13**, 1035–1041.
27 Bailly,C., Dongchul,S., Waring,M.J. and Chaires,J.B. (1998) *Biochemistry*, **37**, 1033–1045.
28 Howard,F.B., Chen,C.W., Cohen,J.S. and Miles,H.T. (1984) *Biochem. Biophys. Res. Commun.*, **118**, 848–853.
29 Kutyavin,I.V., Rhinehart,R.L., Lukhtanov,E.A., Gorn,V.V., Meyer,R.B.,Jr and Gamper,H.B.,Jr (1996) *Biochemistry*, **35**, 11170–11176.
30 Lutz,M.J., Held,H.A., Hottiger,M., Hübscher,U. and Benner,S.A. (1996) *Nucleic Acids Res.*, **24**, 1308–1313.
31 Haaima,G., Hansen,H.F., Christensen,L., Dahl,O. and Nielsen,P.E. (1997) *Nucleic Acids Res.*, **25**, 4639–4643.
32 Kamath-Loeb,A., Hizi,A., Kasai,H. and Loeb,L.A. (1997) *J. Biol. Chem.*, **272**, 5892–5898.
33 Horlacher,J., Hottiger,M., Podust,V.N., Hübscher,U. and Benner,S.A. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 6329–6333.
34 Schick,C. and Martin,C.T. (1993) *Biochemistry*, **32**, 4275–4280.
35 Schick,C. and Martin,C.T. (1995) *Biochemistry*, **34**, 666–672.
36 Bailly,C. and Waring,M.J. (1995) *Nucleic Acids Res.*, **23**, 885–892.
37 Coll,M., Wang,A.H., van der Marel,G.A., van Boom,J.H. and Rich,A. (1986) *J. Biomol. Struct. Dyn.*, **4**, 157–172.
38 Schneider,B., Ginell,S.L., Jones,R., Gaffney,B. and Berman,H.M. (1992) *Biochemistry*, **31**, 9622–9628.
39 Kagawa,T.F., Howell,M.L., Tseng,K. and Ho,P.S. (1993) *Nucleic Acids Res.*, **21**, 5978–5986.
40 Howard,F.B., Limm,W. and Miles,H.T. (1985) *Biochemistry*, **24**, 5033–5039.
41 Borah,B., Cohen,J.S., Howard,F.B. and Miles,H.T. (1985) *Biochemistry*, **24**, 7456–7462.
42 Garriga,P., Garcia-Quintana,D., Sagi,J. and Manyosa,J. (1993) *Biochemistry*, **32**, 1067–1071.
43 Vorlickova,M., Sagi,J., Szabolcs,A., Ebinger,K., Fellegvari,I. and Kypr,J. (1993) *J. Biomol. Struct. Dyn.*, **10**, 681–6692.
44 Kypr,J., Sagi,J., Szakonyi,E., Ebinger,K., Panazova,H., Chladkova,J. and Vorlickova,M. (1994) *Biochemistry*, **33**, 3801–3806.
45 Chazin,W., Rance,M., Chollet,A. and Leupin,W. (1991) *Nucleic Acids Res.*, **19**, 5507–5513.
46 Nielsen,P.E., Møllegaard,N.E. and Jeppesen,C. (1990) *Nucleic Acids Res.*, **18**, 3847–3851.
47 Bailly,C., Mollegaard,N.E., Nielsen,P.E. and Waring,M.J. (1995) *EMBO J.*, **14**, 2121–2131.
48 Diekmann,S., von Kitzing,E., McLaughlin,L.W., Ott,J. and Eckstein,F. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 8257–8261.
49 Mollegaard,N.E., Bailly,C., Waring,M.J. and Nielsen,P.E. (1997) *Nucleic Acids Res.*, **25**, 3497–3502.
50 Strobel,S.A. and Shetty,K. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 2903–2908.
51 Strobel,S.A., Cech,T.R., Usman,N. and Beigelman,L. (1994) *Biochemistry*, **33**, 13824–13835.
52 Strobel,S.A. and Cech,T.R. (1996) *Biochemistry*, **35**, 1201–1211.
53 Kim,Y., Grable,J.C., Love,R., Greene,P.J. and Rosenberg,J.M. (1990) *Science*, **249**, 1307–1309.
54 Brennan,C.A., Van Cleve,M.D. and Gumport,R.I. (1986) *J. Biol. Chem.*, **261**, 7270–7278.
55 McLaughlin,L.W., Benseler,F., Graeser,E., Piel,N. and Scholtissek,S. (1987) *Biochemistry*, **31**, 7238–7245.
56 Aiken,C.R., McLaughlin,L.W. and Gumport,R.I. (1991) *J. Biol. Chem.*, **266**, 19070–19078.
57 Szekeres,M. and Matveyev,A.V. (1987) *FEBS Lett.*, **222**, 89–94.
58 Brennan,C.A., Van Cleve,M.D. and Gumport,R.I. (1986) *J. Biol. Chem.*, **261**, 7279–7286.
59 Thielking,V., Dubois,S., Eritja,R. and Guschlbauer,W. (1997) *Biol. Chem.*, **378**, 407–415.
60 Finkel,S.E. and Johnson,R.C. (1992) *Mol. Microbiol.*, **6**, 3257–3265.
61 Bailly,C., Waring,M.J. and Travers,A.A. (1995) *J. Mol. Biol.*, **253**, 1–7.
62 Bailly,C., Payet,D., Travers,A.A and Waring,M.J. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 13623–13628.
63 Wang,S., Cosstick,R., Gardner,J.F. and Gumport,R.I. (1995) *Biochemistry*, **34**, 13082–13090.
64 Sibghat-Ullah, Xu,Y.-Z. and Day,R.S. (1995) *Biochemistry*, **34**, 7438–7442.
65 Sibghat-Ullah, Gallinari,P., Xu,Y.-Z., Goodman,M.F., Bloom,L.B., Jiricny,J. and Day,R.S. (1996) *Biochemistry*, **35**, 12926–12932.
66 Cerami,A., Reich,E., Ward,D.C. and Goldberg,I.H. (1967) *Proc. Natl Acad. Sci. USA*, **57**, 1036–1042.
67 Sugiyama,H., Fujiwara,T., Kawabata,H., Saito,I., Hirayama,N. and Yoda,N. (1990) *Nucleic Acids Res. Symp. Ser.*, **19**, 55–56.
68 Gao,Y.G., Liaw,Y.C., Li,Y.K., van der Marel,G.A., van Boom,J.H. and Wang,A.H. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 4845–4849.
69 Bailly,C. and Waring,M.J. (1995) *J. Am. Chem. Soc.*, **117**, 7311–7316.
70 Waring,M.J. and Bailly,C. (1997) *J. Mol. Recognition*, **10**, 121–127.
71 Bailly,C., Marchand,C. and Waring,M.J. (1993) *J. Am. Chem. Soc.*, **115**, 3784–3785.
72 Bailly,C. and Waring,M.J. (1998) *Biochem. J.*, **330**, 81–87.
73 Gallego,J., Ortiz,A.R. and Gago,F. (1993) *J. Med. Chem.*, **36**, 1548–1561.
74 Bailly,C., Braña,M. and Waring,M.J. (1996) *Eur. J. Biochem.*, **240**, 195–208.
75 Gallego,J., Luque,F.J., Orozco,M., Burgos,C., Alvares-Builla,J., Rodrigo,M.M. and Gago,F. (1994) *J. Med. Chem.*, **37**, 1602–1609.
76 Gallego,J., Ortiz,A.R., de Pascual-Teresa,B. and Gago,F. (1997) *J. Comput. Aided Mol. Des.*, **11**, 114–128.
77 Jiang,S.P., Raghunathan,G., Ting,K.L. and Jernigan,R.L. (1994) *J. Biomol. Struct. Dyn.*, **12**, 367–382.
78 Berthet,N., Constant,J.F., Demeunynck,M., Michon,P. and Lhomme,J. (1997) *J. Med. Chem.*, **40**, 3346–3352.
79 Murray,V. (1987) *Mutat. Res.*, **177**, 189–199.
80 Hill,F., Williams,D.M., Loakes,D. and Brown,D.M. (1998) *Nucleic Acids Res.*, **26**, 1144–1149.
81 Dutton,F.L.,Jr and Chovnick,A. (1990) *Mol. Gen. Genet.*, **220**, 172–176.

The Purine 2-Amino Group as a Critical Recognition Element for Specific DNA Cleavage by Bleomycin and Calicheamicin

Christian Bailly† and Michael J. Waring*

Contribution from the Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, England

Received January 3, 1995®

Abstract: The influence of the 2-amino group of guanine on antibiotic-mediated cleavage has been studied using DNA in which that group has been removed from guanine, added to adenine, or both. A homologous series of 160 base pair fragments of DNA containing inosine and/or 2,6-diaminopurine residues in place of guanosine and/or adenine residues respectively were synthesized by the polymerase chain reaction and subjected to sequence-specific cleavage by the iron−bleomycin complex or calicheamicin γ_1^I. Although the 2-amino group is not absolutely required it constitutes a key structural element which directs sequence-specific cleavage of DNA. For bleomycin, relocating it created new cleavage sites at pyrimidine residues lying 3′ to 2,6-diaminopurine residues. For calicheamicin, the presence of a purine 2-amino group adjacent to the cutting site potentiated the cleavage reaction. Sequence recognition by bleomycin seems to rely on direct interaction with guanine whereas DNA conformation/flexibility appears more important for calicheamicin.

A large number of potent anticancer drugs owe their efficacy to their ability to promote DNA degradation. Such is the case for the bleomycins and the calicheamicins (Figure 1), two families of glycoconjugate antibiotics capable of inducing DNA strand breakage via free radical mediated mechanisms.[1−3] Cleavage of DNA by bleomycin is dependent on the participation of a redox-active metal ion and a source of oxygen[4−6] whereas calicheamicin and other enediyne compounds require a thiol cofactor.[7−9] In both cases the cleavage reaction proceeds via an attack on deoxyribose by highly reactive species produced upon chemical activation of the drug, but the molecular mechanisms of free radical generation and of DNA cleavage are different. The bleomycin−Fe^{II} complex combines with O_2 to produce a reactive oxygenated metallobleomycin species which is capable of abstracting a hydrogen atom from the deoxyribose ring.[2] Bleomycin generates mainly single strand breaks whereas calicheamicin produces almost exclusively double strand breaks.[10] The chemistry of calicheamicin and other enediyne compounds has been examined in detail in recent years: a reducing agent (e.g. glutathione, dithiothreitol) acts as a nucleophile to initiate the reaction of the trisulfide moiety with the enediyne system which then undergoes a Bergman cyclization reaction leading to a DNA-damaging benzenoid diradical.[3] A feature common to both agents is the high sequence-selectivity of the DNA cleaving reaction: pyrimidine residues 3′ to a guanine residue (i.e. GpC and GpT sequences) are preferentially cleaved by bleomycin[11−15] whereas calicheamicin selectively attacks a pyrimidine residue embedded

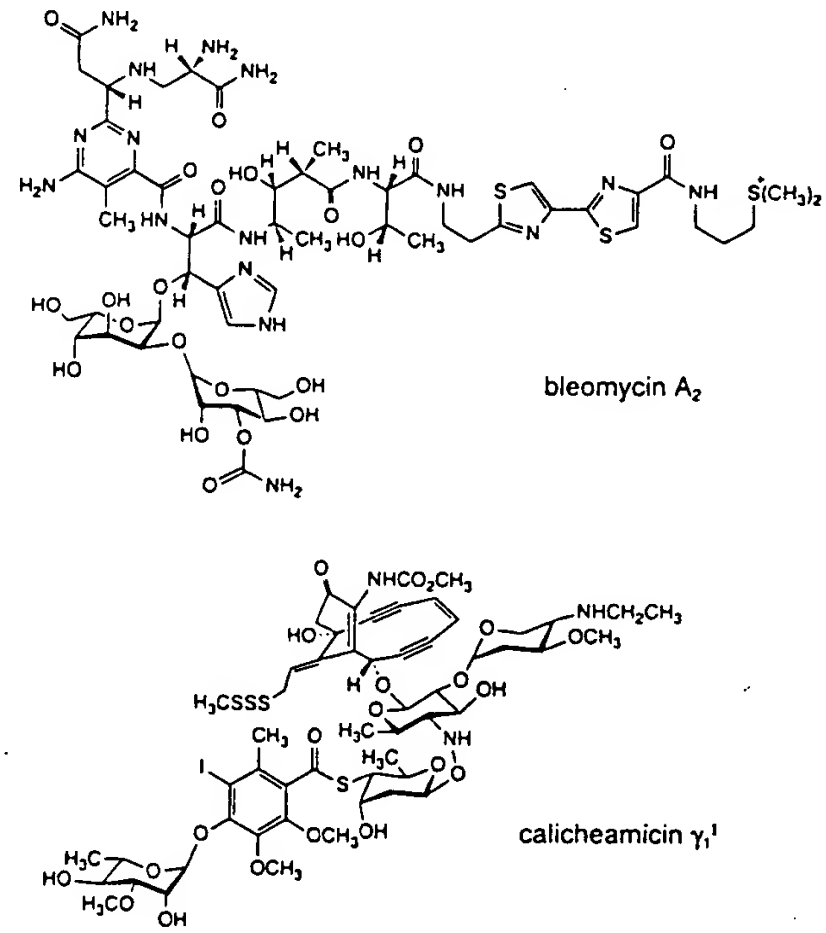


Figure 1. Structures of bleomycin A_2 and calicheamicin γ_1^I.

† Present address: Institut de Recherches sur le Cancer, INSERM Unité 124, Place de Verdun, 59045 Lille cedex, France. FAX (+33) 20 52 70 83.

® Abstract published in *Advance ACS Abstracts,* June 15, 1995.

(1) (1) Dedon, P. C.; Goldberg, I. H. *Chem. Res. Toxicol.* **1992**, *5*, 311−332.

(2) Natrajan, A.; Hecht, S. M. *Molecular Aspects of Anticancer Drug−DNA Interactions*; Neidle, S., Waring, M. J., Eds.; Macmillan: Basingstoke and London, 1994; Vol. 2, pp 197−242.

(3) Ellestad, G. A.; Ding, W.-D. *Molecular Aspects of Anticancer Drug−DNA Interactions*; Neidle, S., Waring, M. J., Eds.; Macmillan: Basingstoke and London, 1994; Vol. 2, pp 130−161.

(4) Hecht, S. M. *Acc. Chem. Res.* **1986**, *19*, 383−391.

(5) Stubbe, J.; Kozarich, J. W. *Chem. Rev.* **1987**, *87*, 1107−1136.

(6) Hecht, S. M. *Bioconjugate Chem.* **1994**, *5*, 513−526.

(7) Nicolaou, K. C.; Dai, W.-M.; Tsay, S.-C.; Estevez, V. A.; Wrasidlo, W. *Science* **1992**, *256*, 1172−1178.

(8) Nicolaou, K. C.; Dai, W.-M. *Angew. Chem., Int. Ed. Engl.* **1991**, *30*, 1387−1530.

(9) Nicolaou, K. C.; Smith, A. L.; Yue, E. W. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5881−5888.

(10) Dedon, P. C.; Salzberg, A. A.; Xu, J. *Biochemistry* **1993**, *32*, 3617−3622.

(11) Mirabelli, C. K.; Ting, A.; Huang, C.-H.; Mong, S.; Crooke, S. T. *Cancer Res.* **1982**, *42*, 2779−2785.

(12) D'Andrea, A. D.; Haseltine, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 3608−3612.

0002-7863/95/1517-7311$09.00/0

in a short homopyrimidine · homopurine tract. Double-stranded lesions produced by calicheamicin γ_1^I (the leading compound in the series) occur with a tetranucleotide specificity mainly at TCCT · AGGA, with other minor sites at short runs of pyrimidines, e.g. TCCG, TCCC, TCTC.[16-18]

The pattern of cleavage is believed to result from prior sequence-selective binding of the antibiotics to DNA.[19-27,28-36] Although the exact modes of binding remain controversial, molecular modeling and experimental studies have suggested that both antibiotics bind within the minor groove of the DNA helix where the 2-amino group of guanine may constitute a critical sequence recognition element. Models have been proposed which involve, for bleomycin, hydrogen bonding between one of its bithiazole nitrogens and the guanosine 2-amino group[37-38] and, for calicheamicin, interaction between the polarizable iodine atom on its benzoate core and the exocyclic amino group of the 5′ guanine in the sequence AGGA.[39,40] The bleomycin model appears consistent with footprinting studies on bithiazole-netropsin hybrid ligands,[41] prompting the inference that interaction of the antibiotic with

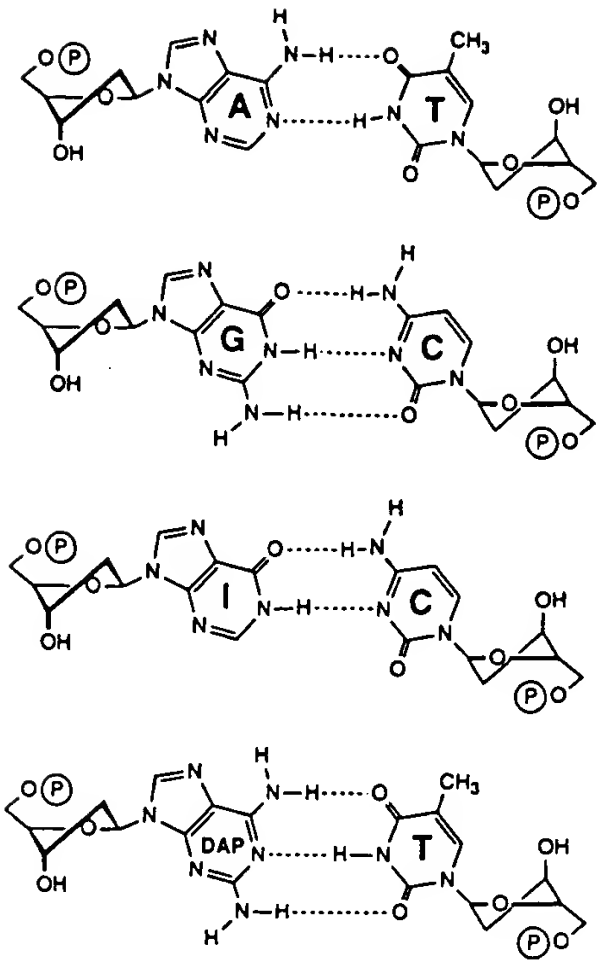


Figure 2. Structures of hydrogen-bonded purine−pyrimidine base pairs. Broken lines represent hydrogen bonds. I represents inosine; DAP represents 2,6-diaminopurine (2-aminoadenine).

the 2-amino group of guanine must somehow be responsible for the sequence specificity of DNA cleavage by the iron−bleomycin complex. By contrast, footprinting studies on calicheamicin γ_1^I suggest that interaction with guanine is likely to be a minor factor in determining the sites of cleavage.[42]

To examine directly the influence of this substituent on DNA cleavage by the two antibiotics we have observed the reaction using DNA in which the purine 2-amino group has been removed, added, or shifted. Removal amounts to converting guanosine nucleotides to inosines (Figure 2). Addition can be accomplished by a modification such that the 2-amino group is present on all purines: this involves replacing the adenines with 2,6-diaminopurines (DAP, Figure 2) while retaining it on guanines. Shifting the 2-amino group can be accomplished by relocating it on to the A · T base pairs (A → DAP substitution) while at the same time removing it from the G · C pairs (G → I substitution). Therefore, by comparing the patterns of cleavage with each of these DNA fragments we can observe the precise effect of the purine 2-amino group on the sequence-specific cleavage of DNA by bleomycin−Fe^{II} complex and calicheamicin.

Results

A series of 160 base pair *tyr*T(A93) DNA fragments[43] containing normal or modified nucleotides were synthesized using the polymerase chain reaction (PCR), labeled on one or another of the complementary strands, and then subjected to a standard cleavage reaction in the presence of each antibiotic.

Cleavage by Bleomycin. The cleavage pattern generated with the bleomycin−Fe^{II} complex varies markedly depending on whether the 2-amino group is present on guanine or adenine

(13) Takeshita, M.; Grollman, A. P.; Ohtsubo, E.; Ohtsubo, H. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 5983−5987.
(14) Fox, K. R.; Grigg, G. W.; Waring, M. J. *Biochem. J.* **1987**, *243*, 847−851.
(15) McLean, M. J.; Dar, A.; Waring, M. J. *J. Mol. Recognit.* **1989**, *1*, 184−192.
(16) Zein, N.; Sinha, A. M.; McGahren, W. J.; Ellestad, G. A. *Science* **1988**, *240*, 1198−1201.
(17) Zein, N.; Poncin, M.; Ramaswamy, R.; Ellestad, G. A. *Science* **1989**, *245*, 697−699.
(18) Walker, S.; Landovitz, R.; Ding, W. D.; Ellestad, G. A.; Kahne, D. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4608−4612.
(19) Sugiyama, H.; Kilkuskie, R. E.; Chang, L.-H.; Ma, L.-T.; Hecht, S. M.; van der Marel, G. A.; van Boom, J. H. *J. Am. Chem. Soc.* **1986**, *108*, 3852−3854.
(20) Shipley, J. B.; Hecht, S. M. *Chem. Res. Toxicol.* **1988**, *1*, 25−27.
(21) Levy, M. J.; Hecht, S. M. *Biochemistry* **1988**, *27*, 2647−2650.
(22) Chikira, M.; Antholine, W. E.; Petering, D. H. *J. Biol. Chem.* **1989**, *264*, 21478−21480.
(23) Carter, B. J.; Murty, V. S.; Reddy, K. S.; Wang, S.-N.; Hecht, S. M. *J. Biol. Chem.* **1990**, *265*, 4193−4196.
(24) Chikira, M.; Sato, T.; Antholine, W. E.; Petering, D. H. *J. Biol. Chem.* **1991**, *266*, 2859−2863.
(25) Hamamichi, N.; Natrajan, A.; Hecht, S. M. *J. Am. Chem. Soc.* **1992**, *114*, 6278−6291.
(26) Urata, H.; Ueda, Y.; Usami, Y.; Akagi, M. *J. Am. Chem. Soc.* **1993**, *115*, 7135−7138.
(27) Fulmer, P.; Petering, D. H. *Biochemistry* **1994**, *33*, 5319−5327.
(28) De Voss, J. J.; Townsend, C. A. *J. Am. Chem. Soc.* **1990**, *112*, 9669−9670.
(29) Drak, J.; Iwazawa, N.; Danishefsky, S.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 7464−7468.
(30) Ding, W.-D.; Ellestad, G. A. *J. Am. Chem. Soc.* **1991**, *113*, 6617−6620.
(31) Kishikawa, H.; Jiang, Y.-P.; Goodisman, J.; Dabrowiak, J. C. *J. Am. Chem. Soc.* **1991**, *113*, 5434−5440.
(32) Aiyar, J.; Danishefsky, S. J.; Crothers, D. M. *J. Am. Chem. Soc.* **1992**, *114*, 7552−7554.
(33) Nicolaou, K. C.; Tsay, S.-C.; Suzuki, T.; Joyce, G. F. *J. Am. Chem. Soc.* **1992**, *114*, 7555−7557.
(34) Walker, S.; Murnick, J.; Kahne, D. *J. Am. Chem. Soc.* **1993**, *115*, 7954−7961.
(35) Aiyar, J.; Hitchcock, S. A.; Denhart, D.; Liu, K. K. C.; Danishefsky, S. J.; Crothers, D. M. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 855−858.
(36) Chatterjee, M.; Cramer, K. D.; Townsend, C. A. *J. Am. Chem. Soc.* **1994**, *116*, 8819−8820.
(37) Kuwahara, J.; Sugiura, Y. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 2459−2463.
(38) Manderville, R. A.; Ellena, J. F.; Hecht, S. M. *J. Am. Chem. Soc.* **1994**, *116*, 10851−852.
(39) Hawley, R. C.; Kiessling, L. L.; Schreiber, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 1105−1109.
(40) Li, T.; Zeng, Z.; Estevez, V. A.; Baldenius, K. U.; Nicolaou, K. C.; Joyce, G. F. *J. Am. Chem. Soc.* **1994**, *116*, 3709−3715.
(41) Bailly, C.; Colson, P.; Houssier, C.; Houssin, R.; Mrani, D.; Gosselin, G.; Imbach, J. L.; Waring, M. J.; Lown, J. W.; Hénichart, J. P. *Biochemistry* **1992**, *31*, 8349−8362.
(42) Mah, S. C.; Townsend, C. A.; Tullius, T. D. *Biochemistry* **1994**, *33*, 614−621.
(43) Drew, H. R.; Weeks, J. R.; Travers, A. A. *EMBO J.* **1985**, *4*, 1025−1032.

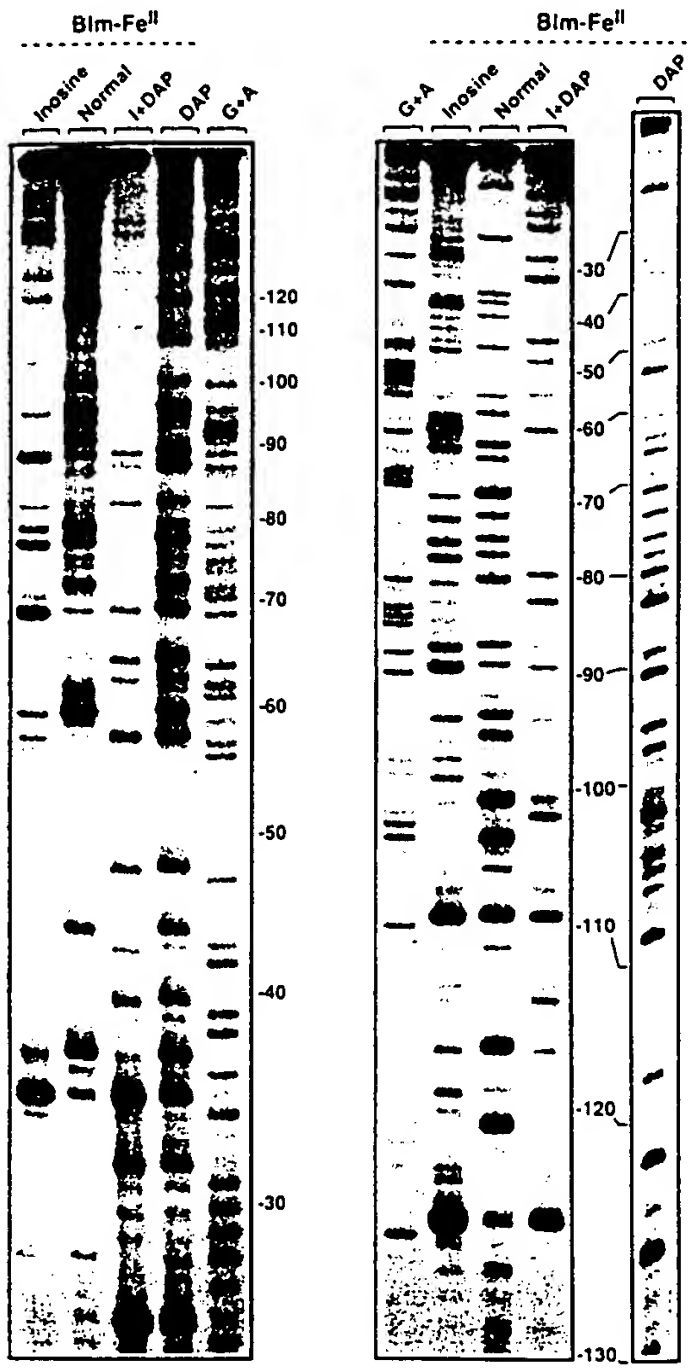


Figure 3. Autoradiographs showing cleavage of normal and modified DNA by the bleomycin−Fe^{II} complex (5 μM). The left and right panels refer respectively to the labeled Watson and Crick strands of *tyr* T-(A93) DNA containing the four natural nucleotides (normal DNA), inosine residues in place of guanosine (Inosine DNA), diaminopurine residues in place of adenine (DAP DNA), or inosine plus DAP residues in place of guanosine and adenine respectively (I+DAP DNA). Chemical identities of the cleavage products were assigned by reference to formic acid−piperidine markers specific for purine residues (lanes G+A), corrected for the expected 1−1.5 band shift due to 5′ end-labeling.[51,52] The G+A track shown in the right pair of panels was determined with the I+DAP-substituted DNA so it is strictly an I+DAP track; it is characterized by much stronger bands at the DAP residues than the inosines, which helps to confirm the correctness of the sequence and the faithful incorporation of I in place of G and DAP in place of A. The scales on the sides of the autoradiographs correspond to the standard numbering of the *tyr* T(A93) sequence as represented in Figure 4.

residues or both (Figure 3). The differences between normal and modified DNA are summarized in the histograms shown in Figure 4. With normal DNA, the iron−bleomycin complex cuts most strongly at GpC and GpT sequences (the underlined nucleotide indicates the reactive site). Some ApC and GpA sites, plus occasional ApT steps, are cleaved as well. When every purine residue bears a 2-amino group (DAP DNA) cleavage can be observed at almost all GpC, GpT, DAPpC, and DAPpT sites, i.e. at nearly all 5′-purine−pyrimidine (RpY) dinucleotide steps. Yet DNA fragments completely lacking the 2-amino group remain susceptible to cleavage by bleomycin. Inosine DNA is cut best at 5′-ApC sites, with weaker cleavage at certain IpC and ApT sequences. With Inosine DNA the majority of the strong cutting sites have a T residue at position 2 so it seems that 5′-TRYR sequences, particularly TACI, provide the preferred sites for cleavage by the antibiotic.[44] With the doubly substituted DNA containing I·C and DAP·T base pairs, the cleavage occurs principally at DAPpC and DAPpT sites. Occasional IpC sites are weakly cut by the iron−bleomycin complex.

Cleavage by Calicheamicin. Calicheamicin γ_1^I cuts normal DNA at a restricted number of sites such as 5′-TTCA, TTCT, TTAC, TCCC, and TTTT (Figure 5). As expected, the cuts appear displaced asymmetrically toward the 3′ ends of the complementary strands (Figure 6) due to binding of the antibiotic in the minor groove. Calicheamicin γ_1^I is known to cleave normal DNA via a pair of cuts separated by 2−3 nucleotides in a 3′ direction such that at least one of those cuts occurs preferentially at certain sequences.[28−36] Normal DNA is generally less well cleaved than modified DNAs under identical conditions. The difference is emphasized in Figure 7 which shows the results of a comparative concentration-dependence study using normal and doubly-substituted DNAs. Several sites of reaction with the two DNAs occur in much the same places but the intensity at a given antibiotic concentration is much stronger with the I+DAP DNA. This may in part reflect the fact that the readable portion of the *tyr*T(A93) DNA autoradiogram lacks certain calicheamicin-favored sequences such as 5′-TCCT, GCCT, TCCG, or TCTC. Figure 5 reveals additionally that cleavage of inosine DNA appears much more uniform: as well as strong reaction at a few canonical sites such as TTTT or TTCA there is significant non-specific cleavage indicating that removal of the 2-amino group is directly responsible for diminished sequence recognition by the enediyne compound. The results obtained with DAP and I+DAP DNA provide clues as regards the need for a calicheamicin−guanine interaction to promote a high level of DNA breakage. Cleavage at the underlined pyrimidine residue in the sequences TTCA·TGAA and TTCT·AGAA (positions 45 and 52, respectively) is abolished when the adjacent G·C pair is replaced by an I·C pair. Conversely, there are numerous sequences cleaved weakly or not at all in normal DNA which become exquisitely susceptible to calicheamicin attack when a 2-amino group is introduced on to the adjacent base pair. Such is the case at positions 33, 113, 44, and 60 (TTAC·GTAA, CCTT·AAGG, GTTC·GAAC, and TAAC·GTTA, respectively). At the latter three sites the I+DAP DNA is strongly cleaved whereas the DAP DNA remains insensitive, as observed with normal DNA. It is also noteworthy that although calicheamicin γ_1^I can cut homopyrimidine·homopurine sequences lacking a G·C base pair (such as those at positions 49 and 67, TTTT·AAAA and TTTA·TAAA, respectively) the extent of cleavage at such sites is considerably enhanced when a 2-amino group is introduced on to the purine residues of the complementary strand (A → DAP substitution).

Discussion

These results lead to the following conclusions: (i) Adding a 2-amino group on to adenine residues (A → DAP substitution) is sufficient to create new cleavage sites for both antibiotics. (ii) Removing the 2-amino group from guanine residues (G →

(44) The base adjacent to the cleaved dinucleotide can markedly affect the extent of damage. A pyrimidine on the 5′ side increases cleavage efficiency while a purine decreases cleavage. The base on the 3′ side of the dinucleotide can also modify the cleavage intensity to a lesser extent.[45]

(45) Murray, V.; Tan, L.; Matthews, J.; Martin, R. F. *J. Biol. Chem.* **1988**, *263*, 12854−12859.

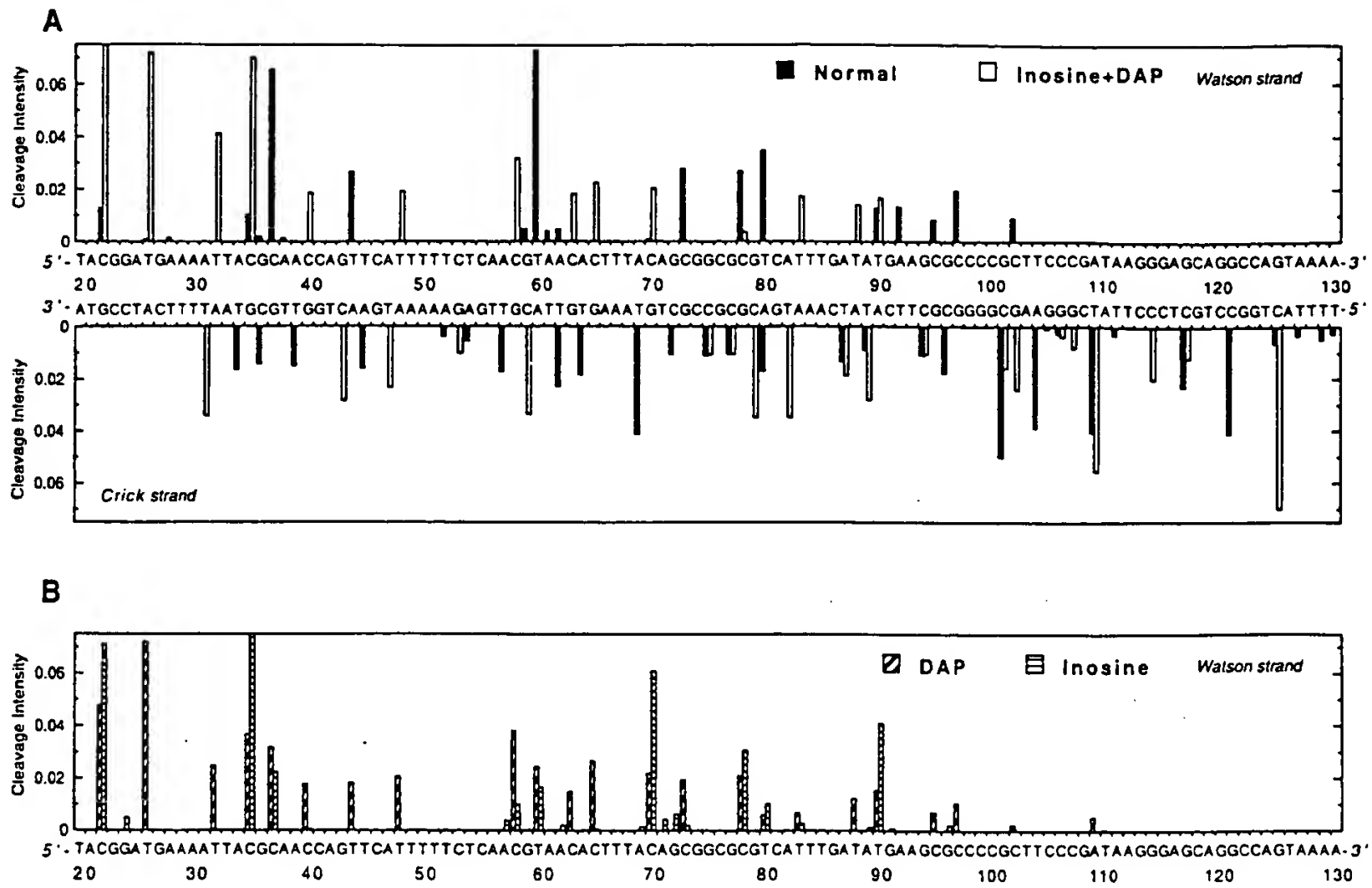


Figure 4. Susceptibility of normal and modified DNA to cleavage by bleomycin. In the modified nucleic acids adenine and/or guanosine residues are replaced by diaminopurine and/or inosine residues, respectively. The relative cleavage intensity (in arbitrary units) at a given bond is expressed as a fraction of the total cleavage of all the phosphodiester bonds within the sequence. Quantitative analysis was limited to regions where peaks were sufficiently well resolved to permit unambiguous analysis. Data are compiled from quantitative analysis of three sequencing gels (including the gel shown in Figure 3) and must be considered as a set of averaged values.

I substitution) does not abolish but reduces the DNA cleavage by bleomycin and greatly diminishes the sequence-specificity of calicheamicin γ_1^I. Therefore the 2-amino group is a powerful determinant but not absolutely necessary for DNA breakage by both bleomycin and calicheamicin. (iii) Shifting the 2-amino group from guanine to adenine residues (A → DAP plus G → I substitutions) produces a complete redistribution of the drug-mediated cleavage sites, showing that both antibiotics are extremely sensitive to the relocation of the purine 2-amino group in DNA.

Thus the 2-amino group of guanine constitutes a key structural element in the mechanism of DNA cleavage by the iron–bleomycin complex as well as by calicheamicin γ_1^I. The results strongly suggest that upon binding to specific sequences the antibiotic molecule engages in contact with the 2-amino group of guanine exposed in the minor groove, consistent with proposed models.[37-41] In particular, the results reported here agree fully with recent NMR studies on the interaction of bleomycin–Zn and bleomycin–Co complexes with the oligonucleotides $d(CGCTAGCG)_2$ and $d(CCAGGCCTGG)_2$, respectively.[38,46] In both cases, the bleomycin molecule is engaged in direct contact with the 2-amino group of guanine, either via one of its bithiazole ring nitrogens (Zn–bleomycin) or via the methyl group of the pyrimidinyl moiety (Co–bleomycin). For calicheamicin, our observations are entirely consistent with the recent NMR experiments on the interaction between d(GGAGCGC)·d(GCGCTCC) and the neocarzinostatin chromophore, another enediyne system, showing that the guanine exocyclic amino group plays a critical role in stabilizing the binding of this drug to double-stranded DNA.[47,48] However, it is clear that other aspects of DNA structure contribute to sequence recognition, especially by calicheamicin. Base substitutions remote from the cutting site can significantly affect the extent of cleavage by bleomycin,[45] and it has been suggested that the flexibility/deformability of the pyrimidine strand is exploited by calicheamicin γ_1^I to distort the DNA structure so that the drug can fit within the minor groove.[49-52] The fact that the inosine-containing DNA, expected to be a good deal more flexible than normal DNA, provides an acceptable substrate for bleomycin and calicheamicin γ_1^I cleavage is in accord with these ideas. Our data vindicate the hypothesis[49-53] that both DNA structure and interaction with guanine are involved in determining sequence-specific cleavage of DNA by bleomycin and calicheamicin.

Experimental Section

Antibiotics, Chemicals, and Biochemicals. Blenoxane (a mixture of 60% bleomycin A_2, 30% bleomycin B_2, and 10% other bleomycins)

(46) Wu, W.; Vanderwall, D. E.; Stubbe, J.; Kozarich, J. W.; Turner, C. J. *J. Am. Chem. Soc.* **1994**, *116*, 10843–10844.

(47) Gao, X.; Stassinopoulos, A.; Rice, J. S.; Goldberg, I. H. *Biochemistry* **1995**, *34*, 40–49.

(48) Sugiyama, H.; Fujiwara, T.; Kawabata, H.; Yoda, N.; Hirayama, N.; Saito, I. *J. Am. Chem. Soc.* **1992**, *114*, 5573–5578.

(49) Uesugi, M.; Sugiura, Y. *Biochemistry* **1993**, *32*, 4622–4627.

(50) Walker, S. L.; Andreotti, A. H.; Kahne, D. E. *Tetrahedron* **1994**, *50*, 1351–1360.

(51) Mah, S. C.; Price, M. A.; Townsend, C. A.; Tullius, T. D. *Tetrahedron* **1994**, *50*, 1361–1378.

(52) Krishnamurthy, G.; Brenowitz, M. D.; Ellestad, G. A. *Biochemistry* **1995**, *34*, 1001–1010.

(53) Nightingale, K. P.; Fox, K. R. *Nucleic Acids Res.* **1993**, *21*, 2549–2555.

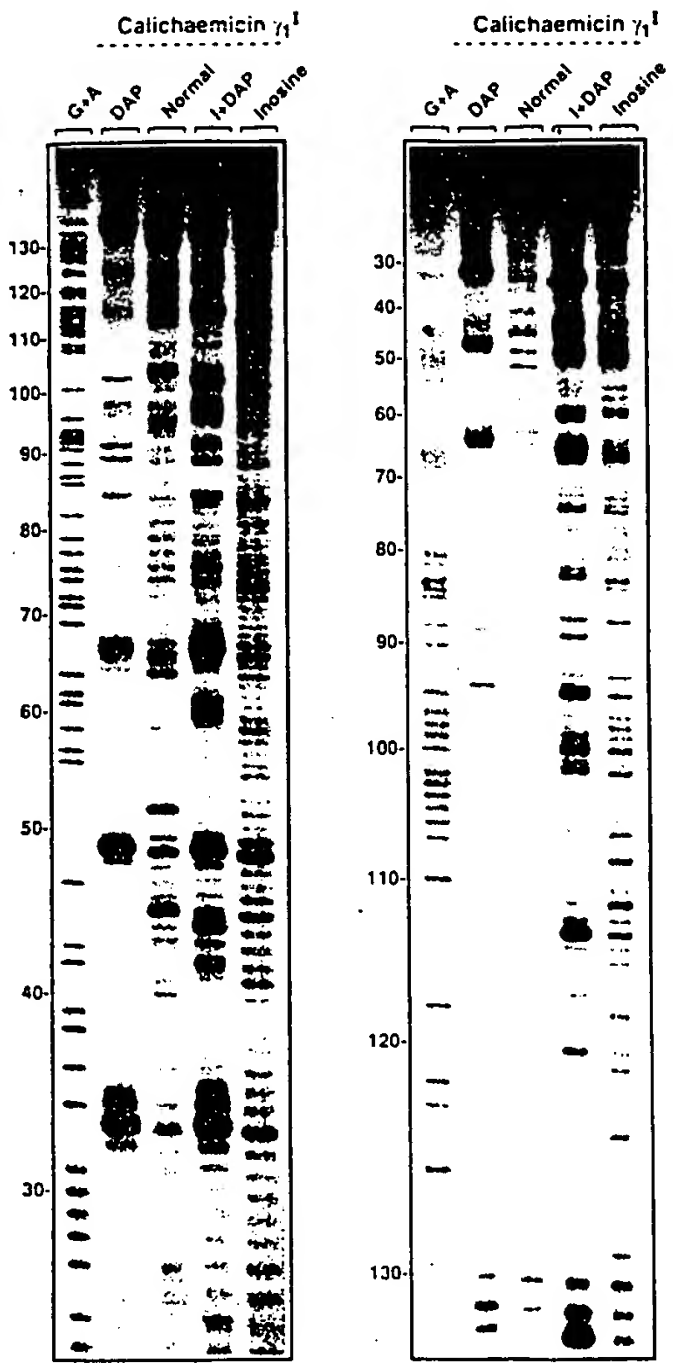


Figure 5. Cleavage of normal and modified DNA by calicheamicin γ_1^I (0.35 μg/mL). The left and right panels refer to the 5′-end labeled Watson and Crick strands of *tyr*T(A93) DNA, respectively. Other details as for Figure 4.

was obtained from Lunbeck. Stock solutions were prepared in 10 mM Tris−HCl buffer containing 10 mM NaCl (pH 7.0), divided into 250 μM aliquots, and stored at −20 °C. Calicheamicin γ_1^I from Cyanamid was dissolved in ethanol to furnish a stock solution at 1.75 μg/mL. Ammonium persulfate, tris base, acrylamide, bis-acrylamide, ultrapure urea, boric acid, tetramethylethylenediamine, and dimethyl sulfate were from BDH. Formic acid, piperidine, and formamide were from Aldrich. Photographic requisites were from Kodak. Bromophenol blue and xylene cyanol were from Serva. Unlabeled deoxynucleoside triphosphates, including dITP, were purchased from Pharmacia. The nucleoside triphosphate labeled with [^{32}P] (γ-ATP; 6000 Ci/mmol) was obtained from NEN Dupont. 2,6-Diaminopurine deoxyribonucleoside triphosphate was obtained by phosphorylation of the corresponding nucleoside (Sigma) according to published procedures.[54,55] Restriction endonucleases *EcoR*I and *Ava*I (Boehringer), *Taq* polymerase (Promega), DNase I (Sigma), and T4 polynucleotide kinase (Pharmacia) were used according to the supplier's recommended protocol in the activity buffer provided. The primers, 5′-AATTCCGGTTACCTTTAATC and 5′-TCGGGAACCCCCACCACGGG having a 5′-OH or 5′-NH_2 terminal group, were synthesized at the Laboratory of Molecular Biology, Medical Research Council, Cambridge. Checks were carried out to ensure that the primers blocked with a 5′-NH_2 group were free from contaminants and not labeled by the kinase. All other chemicals were analytical grade reagents, and all solutions were prepared using doubly deionized, Millipore filtered water.

Preparation, Purification, and Labeling of DNA Fragments Containing Natural and Modified Nucleotides. Plasmid pKMp27[43] was isolated from *E. coli* by a standard sodium dodecyl sulfate-sodium hydroxide lysis procedure and purified by banding in CsCl−ethidium bromide gradients. Ethidium was removed by several 2-propanol extractions followed by exhaustive dialysis against Tris-EDTA buffer. The purified plasmid was then precipitated and resuspended in appropriate buffer prior to digestion by the restriction enzymes. The 160 base pair *tyrT*(A93) fragment used as a template was isolated from the plasmid by digestion with restriction enzymes *Eco*RI and *Ava*I. It is worth mentioning that this template DNA bore a 5′-phosphate due to the action of *Eco*RI and thus only the newly synthesized DNA (with normal or modified nucleotides) can be labeled by the kinase.

(a) Polymerase Chain Reaction (PCR). The protocol used to incorporate inosine and/or 2,6-diaminopurine residues into DNA is comparable to those previously used to incorporate 7-deazapurine or inosine residues with only a few minor modifications.[56−59] PCR reaction mixtures contained 10 ng of *tyrT*(A93) template, 1 μM each of the appropriate pair of primers (one with a 5′-OH and one with a 5′-NH_2 terminal group) required to allow 5′-phosphorylation of the desired strand, 250 μM of each dNTP (dTTP, dCTP plus dATP or dDTP and dGTP or dITP according to the desired DNA), and 5 units of *Taq* polymerase in a volume of 50 μl containing 50 mM KCl, 10 mM Tris−HCl, pH 8.3, 0.1% Triton X-100, and 1.5 mM $MgCl_2$. To prevent unwanted primer-template annealing before the cycles began, the reactions were heated to 60 °C before adding the *Taq* polymerase.[60] Finally, paraffin oil was added to each reaction to prevent evaporation. After an initial denaturing step of 3 min at 94 °C, 20 amplification cycles were performed, with each cycle consisting of the following segments: 94 °C for 1 min, 37 °C for 2 min, and 72 °C for 10 min. After the last cycle, the extension segment was continued for an additional 10 min at 72 °C, followed by a 5-min segment at 55 °C and a 5-min segment at 37 °C. The purpose of these final segments was to maximize annealing of full-length product and to minimize annealing of unused primer to full-length product. The reaction mixtures were then extracted with chloroform to remove the paraffin oil, and parallel reactions were pooled. Several extractions with water-saturated *n*-butanol were performed to reduce the volume prior to loading the samples on to a 6% non-denaturing polyacrylamide gel. After electrophoresis for about 1 h, a thin section of the gel was stained with ethidium bromide so as to locate the band of DNA under UV light. The same band of DNA free of ethidium was excised, crushed, and soaked in elution buffer (500 mM ammonium acetate, 10 mM magnesium acetate) overnight at 37 °C. This suspension was filtered through a Millipore 0.22 μm filter and the DNA was precipitated with ethanol. Following washing with 70% ethanol and vacuum drying of the precipitate, the purified DNA was resuspended in the kinase buffer.

(b) DNA Labeling and Purification. The purified PCR products were 5′-end labeled with [γ-^{32}P]ATP in the presence of T4 polynucleotide kinase according to a standard procedure for labeling blunt-ended DNA fragments.[61] After completion the labeled DNA was again purified by 6% polyacrylamide gel electrophoresis and extracted from the gel as described above. Finally, the labeled DNA was resuspended in 10 mM Tris−HCl buffer at pH 7.0 containing 10 mM NaCl.

Cleavage of DNA by the Bleomycin−Fe^{II} Complex and Calicheamicin γ_1^I. In a typical experiment, the freshly prepared bleomycin−Fe^{II} complex (4 μL) was added to 6 μL of 5′-end labeled DNA (~1 nM) in 10 mM Tris−HCl buffer at pH 7.0 containing 10 mM NaCl. The equimolar bleomycin−Fe complex consisted of 2 μL of a 25 μM

(54) Ludwig, J. *Acta Biochim. Biophys. Acad. Sci., Hung.* **1981**, *16*, 131−133.
(55) Seela, F.; Muth, H.-P.; Röling, A. *Helv. Chim. Acta* **1991**, *74*, 554−564.
(56) Sayers, E. W.; Waring, M. J. *Biochemistry* **1993**, *32*, 9094−9107.
(57) Marchand, C.; Bailly, C.; McLean, M. J.; Moroney, S.; Waring, M. J. *Nucleic Acids Res.* **1992**, *20*, 5601−5606.
(58) Bailly, C.; Marchand, C.; Waring, M. J. *J. Am. Chem. Soc.* **1993**, *115*, 3784−3785.
(59) Bailly, C.; Waring, M. J. *Nucleic Acids Res.* **1995**, *23*, 885−892.
(60) Bloch,W. *Biochemistry* **1991**, *30*, 2735−2747.
(61) Maniatis,T.; Fritsch, E. F.; Sambrook, J. *Molecular Cloning: A Laboratory Manual:* Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1982.

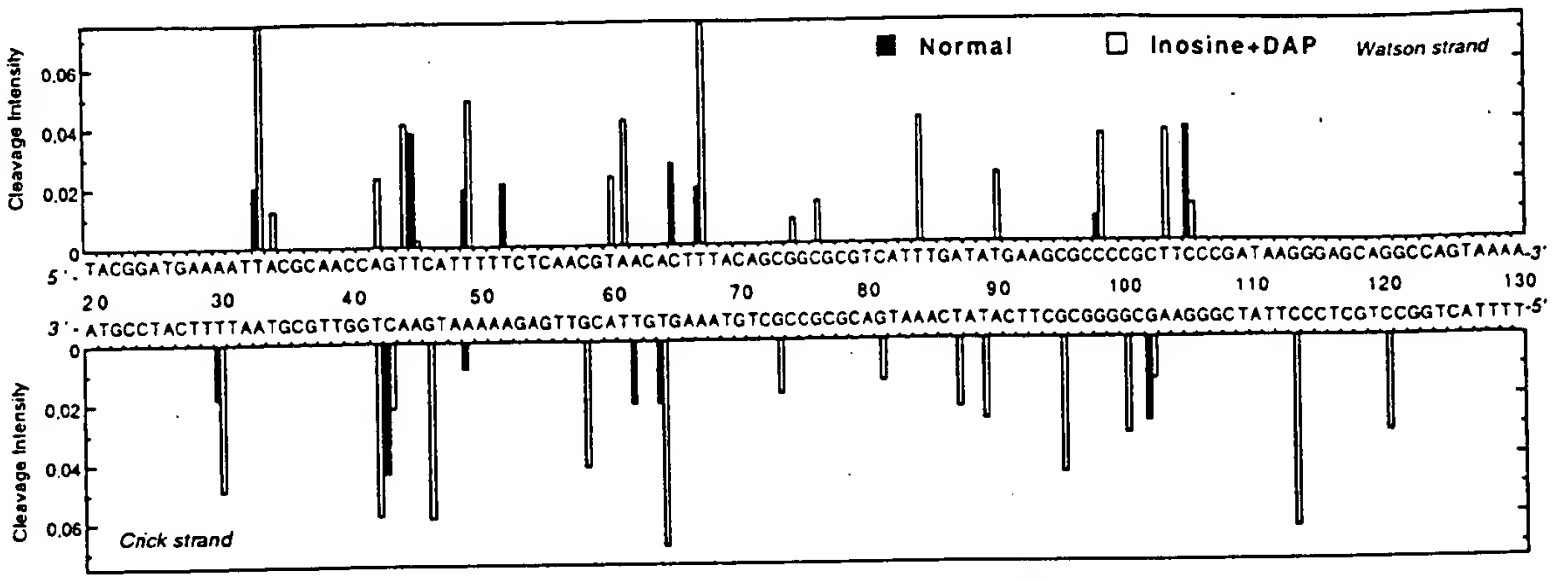


Figure 6. Susceptibility of normal and I+DAP DNA to cleavage by calicheamicin γ_1^I. Details as in Figure 4.

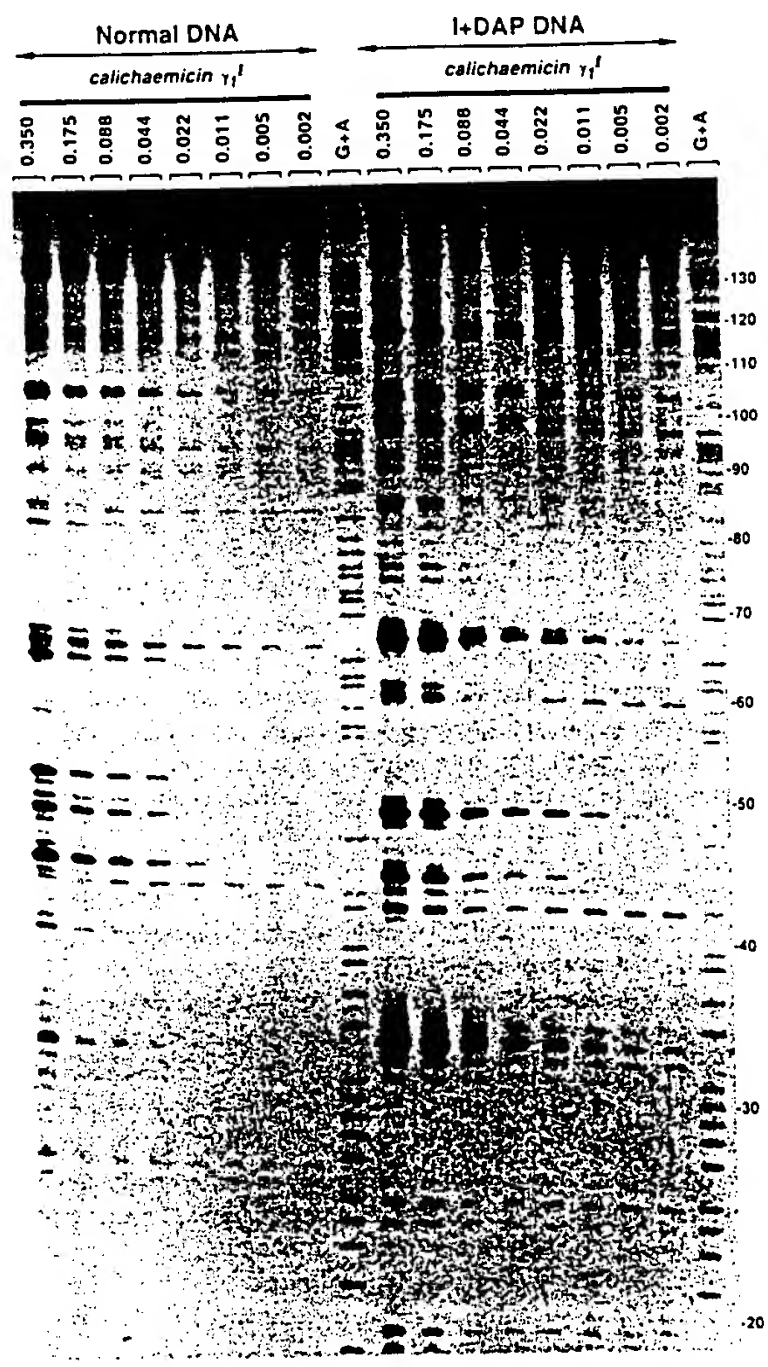


Figure 7. Cleavage of normal and inosine plus DAP-containing DNA in the presence of increasing concentrations of calicheamicin γ_1^I. The reactions were conducted in accordance with the procedure described in the Experimental Section. The calicheamicin concentration in (μg/mL) is indicated at the top of each lane.

solution of blenoxane and 2 μL of a 25 μM solution of $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$ mixed just prior to the experiment. After incubation at room temperature for periods varying from a few seconds to 1 min the cleavage reaction was stopped by freezing. Samples were lyophilized, resuspended in 50 μL of water, and lyophilized again. The cleavage products were resuspended in 4 μL of formamide–dye solution and resolved on a denaturing polyacrylamide gel as described below.

Two microliters of a 1.75-μg/mL stock solution of calicheamicin γ_1^I in ethanol were incubated with 7 μL of DNA (~1 nM) in 10 mM Tris–HCl buffer at pH 7.0 containing 10 mM NaCl. Solutions of calicheamicin were prepared by serially diluting the master stock solution into ethanol. The final ethanol content in the reaction mixture was 20%. An equivalent volume of ethanol was added to the control tubes. The DNA-calicheamicin γ_1^I solutions were equilibrated for 10 min prior to the initiation of the cleavage chemistry. At the chosen time, 1 μL of 1 mM dithiothreitol was added and the reaction allowed to proceed for 5 min at room temperature. After precipitation with ethanol, the DNA sample was resuspended in 4 μL of formamide–dye solution and the products resolved on a denaturing polyacrylamide gel. Samples were heated at 90 °C for 4 min and chilled in ice for 4 min prior to electrophoresis.

Electrophoresis, Autoradiography, and Quantitation by Storage Phosphor Imaging. DNA cleavage products were resolved by polyacrylamide gel electrophoresis under denaturing conditions (0.3 mm thick, 8% acrylamide containing 8 M urea) capable of resolving DNA fragments differing in length by one nucleotide. Electrophoresis was continued until the bromophenol blue marker had run out of the gel (about 2.5 h at 60 W, 1600 V in TBE buffer, BRL sequencer model S2). Gels were soaked in 10% acetic acid for 15 min, transferred to Whatman 3MM paper, dried under vacuum at 80 °C, and examined by autoradiography using either a phosphorimager or X-ray films (Fuji R-X) exposed at −70 °C with an intensifying screen usually for 24 h. For quantitative analysis, a Molecular Dynamics 425E PhosphorImager was used to collect data from storage screens exposed to the dried gels overnight at room temperature. Base line corrected scans were analyzed by integrating all the densities between two selected boundaries using ImageQuant version 3.3 software. Each resolved band was assigned to a particular bond within the *tyr*T(A93) fragment by comparison of its position relative to sequencing standards generated by treatment of the DNA with formic acid followed by piperidine-induced cleavage at the purine residues (G+A track).

Acknowledgment. The authors thank Drs J. Sagi and C. Abell for the preparation of the DAP triphosphate. We thank Dean Gentle for his expert technical assistance. This work was supported by grants to M.J.W. from the Cancer Research Campaign, the Wellcome Trust, and the Association for International Cancer Research. C.B. thanks the INSERM, the ARC, and the Sir Halley Stewart Trust.

JA950024J

Nucleic Acids Research, 1995, Vol. 23, No. 6 885–892

Transferring the purine 2-amino group from guanines to adenines in DNA changes the sequence-specific binding of antibiotics

Christian Bailly+ and Michael J. Waring*

Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, UK

Received January 20, 1995; Revised and Accepted February 9, 1995

ABSTRACT

The proposition that the 2-amino group of guanine plays a critical role in determining how antibiotics recognise their binding sites in DNA has been tested by relocating it, using *tyr*T DNA derivative molecules substituted with inosine plus 2,6-diaminopurine (DAP). Irrespective of their mode of interaction with DNA, such GC-specific antibiotics as actinomycin, echinomycin, mithramycin and chromomycin find new binding sites associated with DAP-containing sequences and are excluded from former canonical sites containing I•C base pairs. The converse is found to be the case for a group of normally AT-selective ligands which bind in the minor groove of the helix, such as netropsin: their preferred sites become shifted to IC-rich clusters. Thus the binding sites of all these antibiotics strictly follow the placement of the purine 2-amino group, which accordingly must serve as both a positive and negative effector. The footprinting profile of the 'threading' intercalator nogalamycin is potentiated in DAP plus inosine-substituted DNA but otherwise remains much the same as seen with natural DNA. The interaction of echinomycin with sites containing the TpDAP step in doubly substituted DNA appears much stronger than its interaction with CpG-containing sites in natural DNA.

INTRODUCTION

Gene targeting via DNA-binding drugs remains a cherished goal of chemotherapy (1–3). If it is to succeed we need a clear understanding of the mechanisms whereby small molecules can recognise and bind to specific nucleotide sequences in DNA. It has long been suspected that the 2-amino group of guanine is the prime element which determines sequence recognition via the minor groove of the helix where most small molecules bind (4). It is the only hydrogen bond donor group exposed there; it impedes access to the floor of the groove and it interferes with the spine of hydration (5). We have tested the importance of the purine 2-amino group by removing it from guanines and relocating it on the adenine residues. Footprinting experiments performed on the *tyr*T DNA fragment thus modified reveal that the binding (recognition) sites for all small sequence-selective ligands tested, including some whose actual specificity is unknown, are profoundly changed.

To accomplish the transfer of the 2-amino group we used the polymerase chain reaction to prepare homologous DNA samples having guanosine nucleotides replaced by inosines, adenine residues replaced by 2,6-diaminopurines (Fig. 1), or both. The modified DNA species, as well as normal DNA prepared by the same route, were then subjected to DNAase I footprinting in the presence of antibiotics known to bind selectively to DNA at particular sequences by disparate mechanisms (Fig. 2). As an example of a well-characterised intercalator we chose actinomycin D and, for comparison, a bis-intercalator of comparable structure and molecular weight, echinomycin: both are cyclic depsipeptides endowed with potent antitumour activity (6). As examples of ligands binding to the minor groove we looked at two antiviral agents, netropsin and distamycin, as well as mithramycin and chromomycin which are also chemotherapeutic drugs used in the treatment of cancer (6). Lastly, we included nogalamycin as a representative of the clinically important anthracycline group of antitumour antibiotics (7).

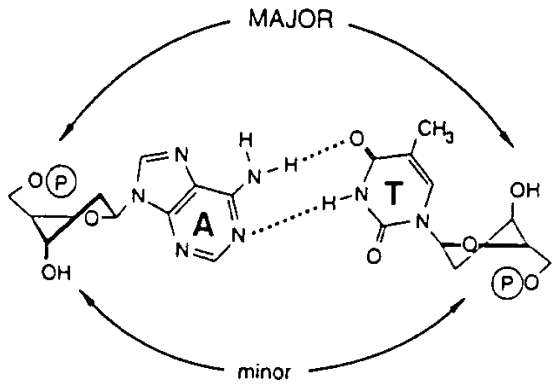
MATERIALS AND METHODS

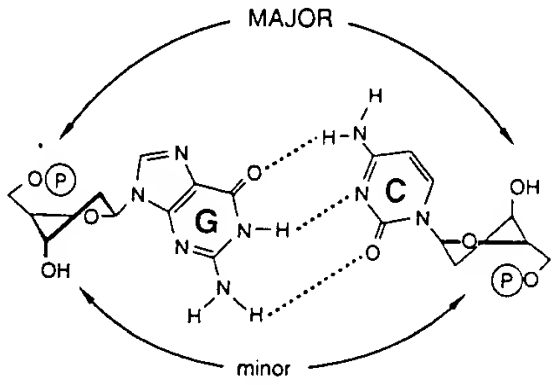
Antibiotics

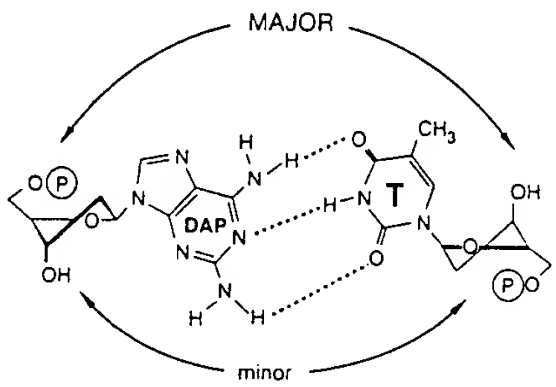
Actinomycin D, nogalamycin, distamycin, chromomycin and mithramycin were purchased from Sigma. Netropsin was from Serva and echinomycin was obtained from Parke-Davis (NJ, USA). Antibiotics were used as supplied without further purification. The tested drugs showed good aqueous solubility except echinomycin which is sparingly soluble in water. Echinomycin was dissolved to a concentration of 100 µM in 10 mM Tris–HCl, pH 7.0, 10 mM NaCl containing 40% (v/v) methanol. The stock solution was diluted to working concentrations with appropriate volumes of 10 mM Tris–HCl, pH 7.0, 10 mM NaCl and methanol so as to yield a final methanol concentration of 10% (v/v) in the footprinting reactions. Under these conditions methanol is known not to affect the nuclease activity (8). Antibiotic concentrations

* To whom correspondence should be addressed

+Present address: Institut de Recherches sur le Cancer, INSERM Unité 124, Place de Verdun, 59045 Lille, France







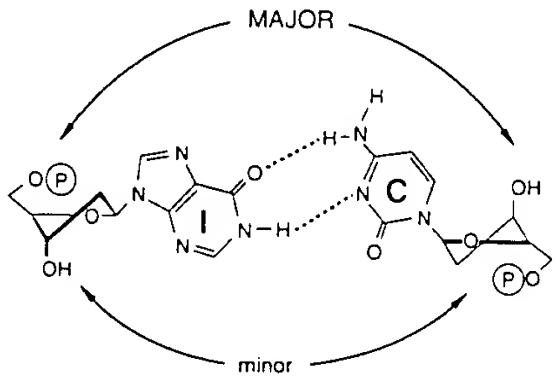


Figure 1. Structures of hydrogen-bonded purine–pyrimidine base pairs. Broken lines represent hydrogen bonds. I represents inosine; DAP represents 2,6-diaminopurine (2-aminoadenine). The major and minor grooves of the helix are indicated.

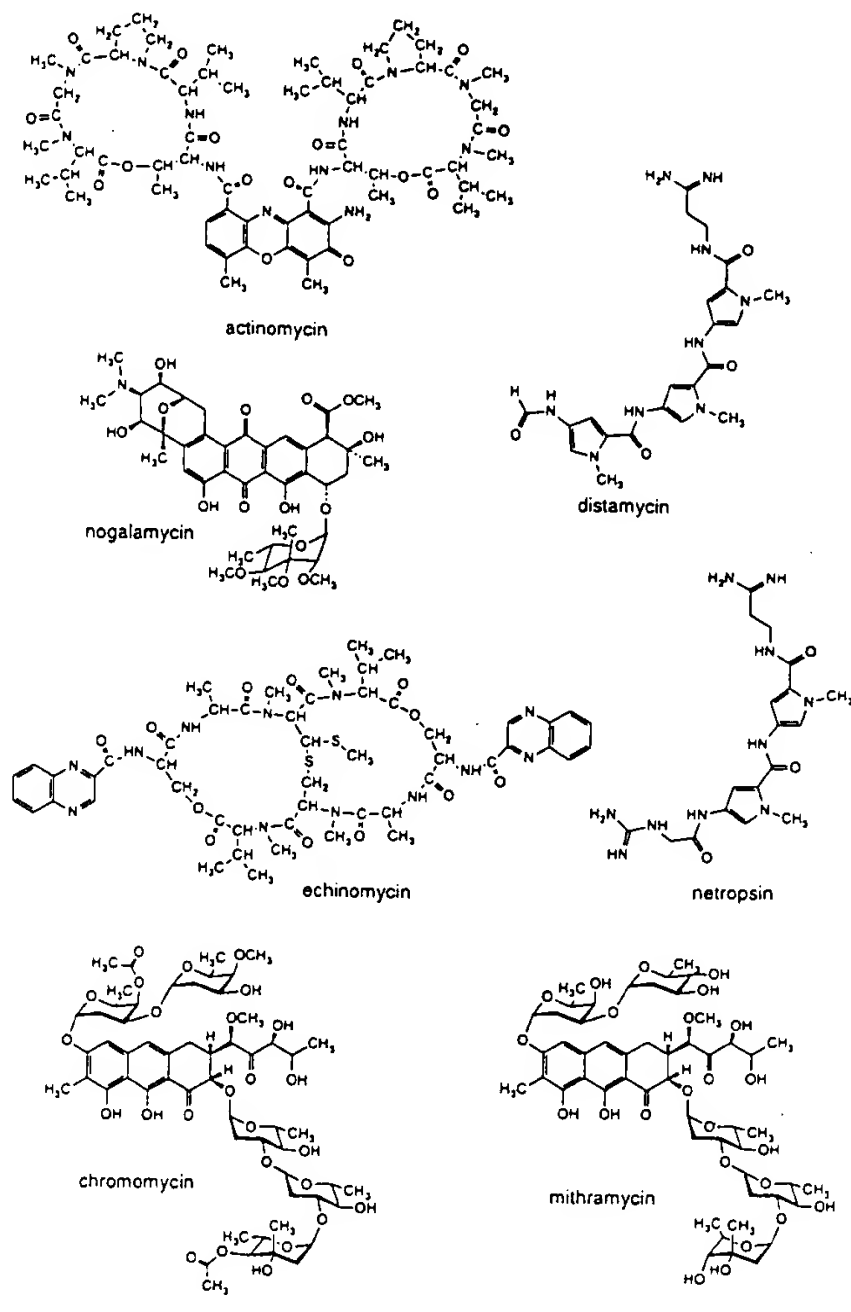


Figure 2. Structures of DNA-binding antibiotics.

were determined spectroscopically in 10 mm pathlength quartz cuvettes through the molar extinction coefficients given in the literature.

Chemicals and biochemicals

Ammonium persulphate, Tris base, acrylamide, bis-acrylamide, ultrapure urea, boric acid, tetramethylethylenediamine and dimethyl sulphate were from BDH. Formic acid, piperidine and formamide were from Aldrich. Photographic requisites were from Kodak. Bromophenol blue and xylene cyanol were from Serva. The nucleoside triphosphate labelled with [γ-^{32}P]ATP was obtained from NEN Dupont. Restriction endonucleases *Eco*RI and *Ava*I (Boehringer), *Taq* polymerase (Promega), DNase I (Sigma) and T4 polynucleotide kinase (Pharmacia) were used according to the supplier's recommended protocol in the activity buffer provided. The primers, 5′-AATTCCGGTTAC-CTTTAATC and 5′-TCGGGAACCCCCACCACGGG having a 5′-OH or 5′-NH_2 terminal group, were obtained from the Laboratory of Molecular Biology, Medical Research Council, Cambridge. Checks were carried out to ensure that the primers blocked with a 5′-NH_2 group were free from contaminants and not labelled by the kinase. All other chemicals were analytical grade reagents, and all solutions were prepared using doubly deionised, Millipore filtered water.

Preparation, purification and labelling of DNA fragments containing natural and modified nucleotides

Plasmid pKMp27 (9) was isolated from *E.coli* by a standard sodium dodecyl sulphate–sodium hydroxide lysis procedure and purified by banding in CsCl–ethidium bromide gradients. Ethidium was removed by several isopropanol extractions followed by exhaustive dialysis against Tris–EDTA buffer. The purified plasmid was then precipitated and resuspended in appropriate buffer prior to digestion by the restriction enzymes. The 160 base

pair *tyrT*(A93) fragment used as a template was isolated from the plasmid by digestion with restriction enzymes *Eco*RI and *Ava*I. It is worth mentioning that this template DNA bore a 5′-phosphate due to the action of *Eco*RI and thus only the newly synthesized DNA (with normal or modified nucleotides) can be labelled by the kinase.

Polymerase chain reaction (PCR). The protocol used to incorporate inosine and/or 2,6-diaminopurine residues into DNA is comparable to those previously used to incorporate 7-deazapurine or inosine residues with only a few minor modifications (10,11). PCR reaction mixtures contained 10 ng of *tyrT*(A93) template, 1 µM each of the appropriate pair of primers (one with a 5′-OH and one with a 5′-NH_2 terminal group) required to allow 5′-phosphorylation of the desired strand, 250 µM of each appropriate dNTP (dTTP, dCTP plus dATP or dDTP and dGTP or dITP according to the desired DNA), and 5 U of *Taq* polymerase in a volume of 50 µl containing 50 mM KCl, 10 mM Tris–HCl, pH 8.3, 0.1% Triton X-100 and 1.5 mM $MgCl_2$. To prevent unwanted primer–template annealing before the cycles began, the reactions were heated to 60°C before adding the *Taq* polymerase (12). Finally, paraffin oil was added to each reaction to prevent evaporation. After an initial denaturing step of 3 min at 94°C, 20 amplification cycles were performed, with each cycle consisting of the following segments: 94°C for 1 min, 37°C for 2 min and 72°C for 10 min. After the last cycle, the extension segment was continued for an additional 10 min at 72°C, followed by a 5 min segment at 55°C and a 5 min segment at 37°C. The purpose of these final segments was to maximize annealing of full-length product and to minimise annealing of unused primer to full-length product. The reaction mixtures were then extracted with chloroform to remove the paraffin oil, and parallel reactions were pooled. Several extractions with water-saturated *n*-butanol were performed to reduce the volume prior to loading the samples on to a 6% non-denaturing polyacrylamide gel. After electrophoresis for ~1 h, a thin section of the gel was stained with ethidium bromide so as to locate the band of DNA under UV light. The same band of DNA free of ethidium was excised, crushed and soaked in elution buffer (500 mM ammonium acetate, 10 mM magnesium acetate) overnight at 37°C. This suspension was filtered through a Millipore 0.22 µm filter and the DNA was precipitated with ethanol. Following washing with 70% ethanol and vacuum drying of the precipitate, the purified DNA was resuspended in the kinase buffer.

DNA labelling and purification. The purified PCR products were 5′ end-labelled with [γ-^{32}P]ATP in the presence of T4 polynucleotide kinase according to a standard procedure for labelling blunt-ended DNA fragments (13). After completion the labelled DNA was again purified by 6% polyacrylamide gel electrophoresis and extracted from the gel as described above. Finally, the labelled DNA was resuspended in 10 mM Tris–HCl, pH 7.0 buffer containing 10 mM NaCl.

DNase I footprinting

DNase I experiments were performed essentially according to the original protocol (8). The digestion of the samples (6 µl) of the labelled DNA fragment dissolved in 10 mM Tris buffer pH 7.0 containing 10 mM NaCl was initiated by the addition of 2 µl of a DNase I solution whose concentration was adjusted to yield a final enzyme concentration of ~0.01 U/ml in the reaction mixture. The extent of digestion was limited to <30% of the starting material so as to minimize the incidence of multiple cuts in any strand ('single-hit' kinetic conditions). Optimal enzyme dilutions were established in preliminary calibration experiments. After 3 min, the digestion was stopped by freeze drying, samples were lyophilized, washed once with 50 µl of water, lyophilized again and then resuspended in 4 µl of an 80% formamide solution containing tracking dyes. Samples were heated at 90°C for 4 min and chilled in ice for 4 min prior to electrophoresis.

Electrophoresis and autoradiography

DNA cleavage products were resolved by polyacrylamide gel electrophoresis under denaturing conditions (0.3 mm thick, 8% acrylamide containing 8 M urea) capable of resolving DNA fragments differing in length by one nucleotide. Electrophoresis was continued until the bromophenol blue marker had run out of the gel (~2.5 h at 60 W, 1600 V in TBE buffer, BRL sequencer model S2). Gels were soaked in 10% acetic acid for 15 min, transferred to Whatman 3MM paper, dried under vacuum at 80°C and subjected to autoradiography at –70°C with an intensifying screen. Exposure times of the X-ray films (Fuji R-X) were adjusted according to the number of counts per lane loaded on each individual gel (usually 24 h).

Quantitation by storage phosphorimaging

A Molecular Dynamics 425E PhosphorImager was used to collect data from storage screens exposed to the dried gels overnight at room temperature (14). Base line-corrected scans were analyzed by integrating all the densities between two selected boundaries using ImageQuant version 3.3 software. Each resolved band was assigned to a particular bond within the *tyr*T(A93) fragment by comparison of its position relative to sequencing standards generated by treatment of the DNA with formic acid followed by piperidine-induced cleavage at the purine residues (G+A track).

RESULTS

Strong footprints were produced by each ligand on the normal and doubly substituted DNAs (Fig. 3). Echinomycin is a bis-intercalator (8,15), and actinomycin a mono-intercalator (16,17), which normally bind to CpG and GpC steps respectively: the footprints for both antibiotics are radically altered by the nucleotide substitution. The same is true for netropsin (18) and distamycin (19,20) which are AT-specific minor groove-binders, as well as for mithramycin and chromomycin (21) which bind in dimeric form to GC-rich sequences within the minor groove (22–24). Even the binding of nogalamycin (25–27), a 'threading' intercalator which interacts with a puzzling variety of sites in natural DNA (28), is changed.

The canonical and newly created binding sites in the two types of DNA can be identified in Figure 4 where the near-inversion of the footprinting pattern for most ligands is clearly evident, especially in the central portion of the sequence extending from position 60 to 100 of the *tyr*T fragment. Irrespective of their mode of binding, the normally GC-selective antibiotics are displaced from their clustered sites lying between positions 70 and 80 in natural DNA, to pick up new sites in the DAP•T-rich sequences either side (panels A, B and D). By contrast, the normally

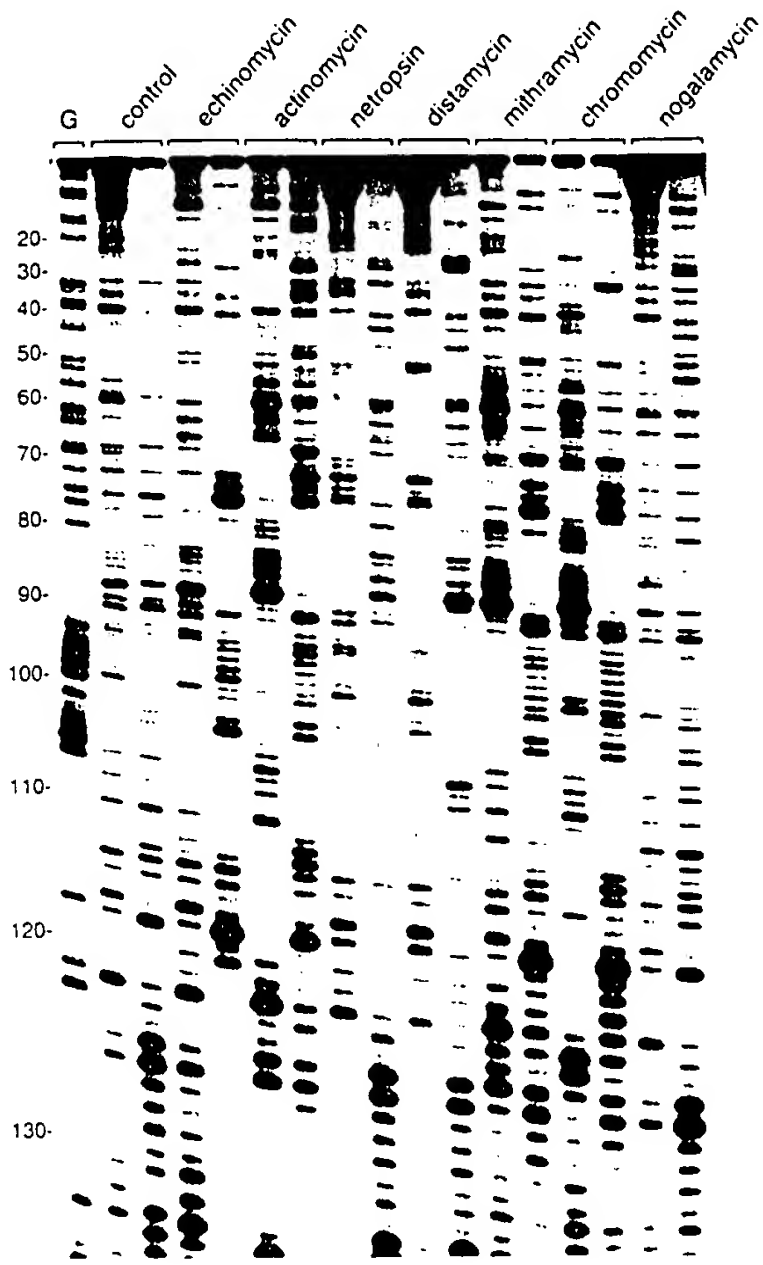


Figure 3. Autoradiograph of a high-resolution denaturing gel showing DNAase I footprinting of antibiotics on the Crick (sense) strand of normal and inosine plus DAP *tyr*T(A93) DNA. Each pair of lanes corresponds to digestion of normal (left) and I + DAP DNA (right). The products of DNAase I digestion were resolved on an 8% polyacrylamide gel containing 8 M urea. Their identities were assigned by reference to the Maxam–Gilbert guanine markers (lane G), taking into account the difference in mobility of the fragments due to the presence or absence of a 3′-phosphate group. Lanes marked control refer to the DNAase I digests of the normal and doubly substituted DNA in the absence of antibiotic. The other lanes contained 20 μM actinomycin, 10 μM netropsin, 10 μM distamycin, 10 μM mithramycin, 10 μM chromomycin or 5 μM nogalamycin. With echinomycin, 20 μM was used to footprint normal DNA and 5 μM for inosine plus DAP-DNA. The scale on the left corresponds to the standard numbering of the *tyr*T(A93) sequence (9) as represented in Figure 5.

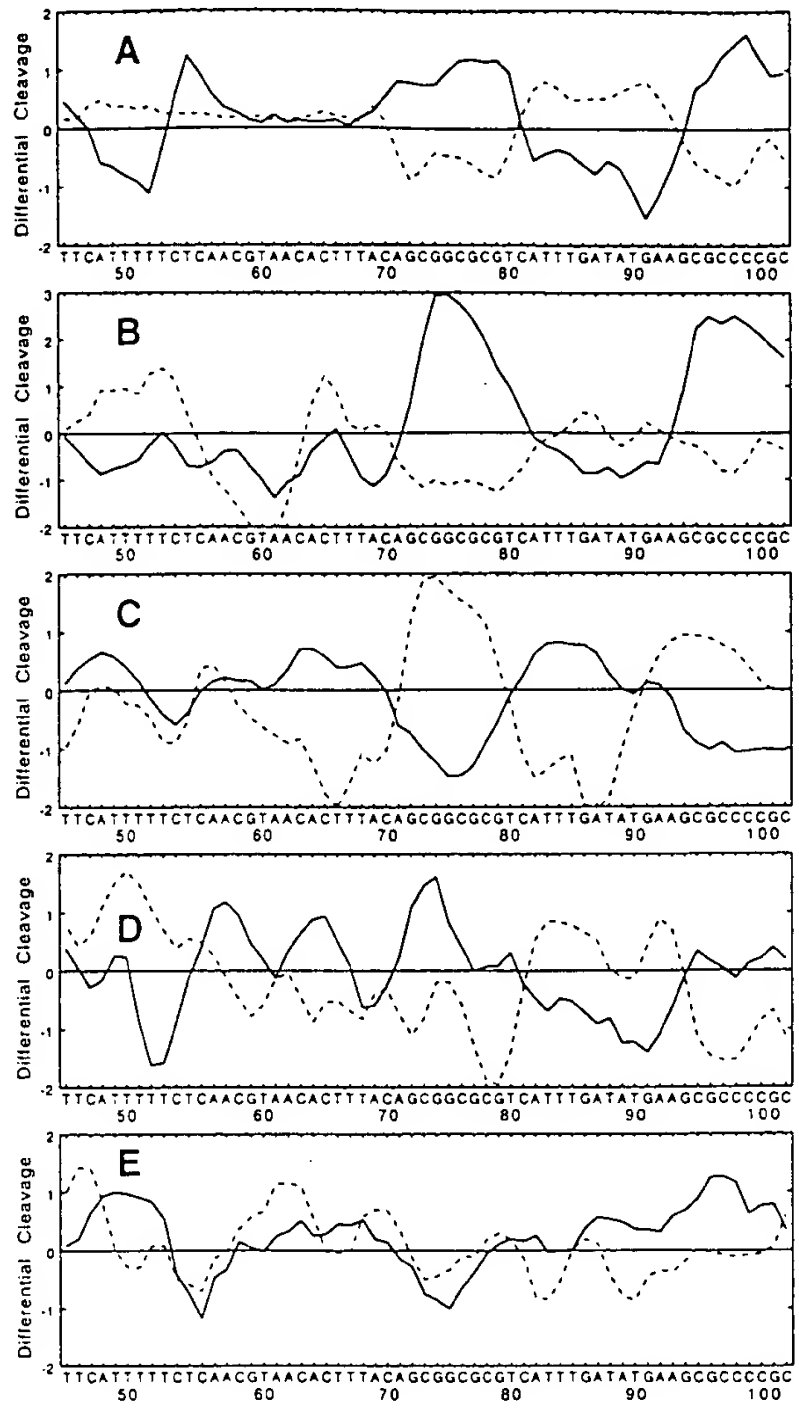


Figure 4. Differential cleavage plots comparing the DNAase I-mediated cleavage of the Watson (antisense) strand of normal and I + DAP *tyr*T(A93) DNA in the presence of actinomycin (**A**), echinomycin (**B**), netropsin (**C**), chromomycin (**D**) and nogalamycin (**E**). Data for distamycin and mithramycin are not included; they were much the same as those shown for netropsin and chromomycin respectively. The plots drawn as continuous lines refer to the modified *tyr*T(A93) DNA fragment containing inosine and DAP residues. The plots indicated by dashed lines refer to normal *tyr*T(A93) DNA. Antibiotic concentrations used to generate these plots were identical to those specified in Figure 3. Positive and negative values correspond, respectively, to enhanced or diminished DNAase I cutting at each internucleotide bond. The values plotted compare the measured probabilities of cleavage expressed in logarithmic units and are smoothed by taking a three-bond running average.

AT-selective netropsin is displaced from its canonical sites in those flanking sequences (positions 60–69 and 82–90) to bind decisively to the intervening IC-rich cluster created in the doubly substituted DNA (panel C). Essentially similar results were obtained with distamycin and with the synthetic minor groove-binding drugs DAPI and berenil (29,30).

The case of echinomycin is examined in greater detail in Figure 5 which shows the complete footprinting pattern on both strands of the *tyr*T(A93) DNA over the whole length of sequence accessible to analysis. As previously reported with this restriction fragment (8) the canonical sites in natural *tyr*T DNA are squarely located around the CpG steps, marked by open rectangles in the illustration. In the doubly inosine plus DAP-substituted DNA every region protected from DNAase I attack occurs at a TpDAP step and all such steps (shown as black rectangles) constitute part of a ligand binding site. The radically changed pattern of binding sites for echinomycin, reflecting its faithful recognition of the displaced 2-amino groups, is emphasised by an examination of the antibiotic concentration-dependence of the footprinting profile (Fig. 6). Over at least two orders of magnitude the footprinting pattern remains firmly locked to the sites surrounding the TpDAP steps at positions 61, 69, 89, 111, 127 and 137 on the Crick strand of the *tyr*T fragment. Only at the TpI step at

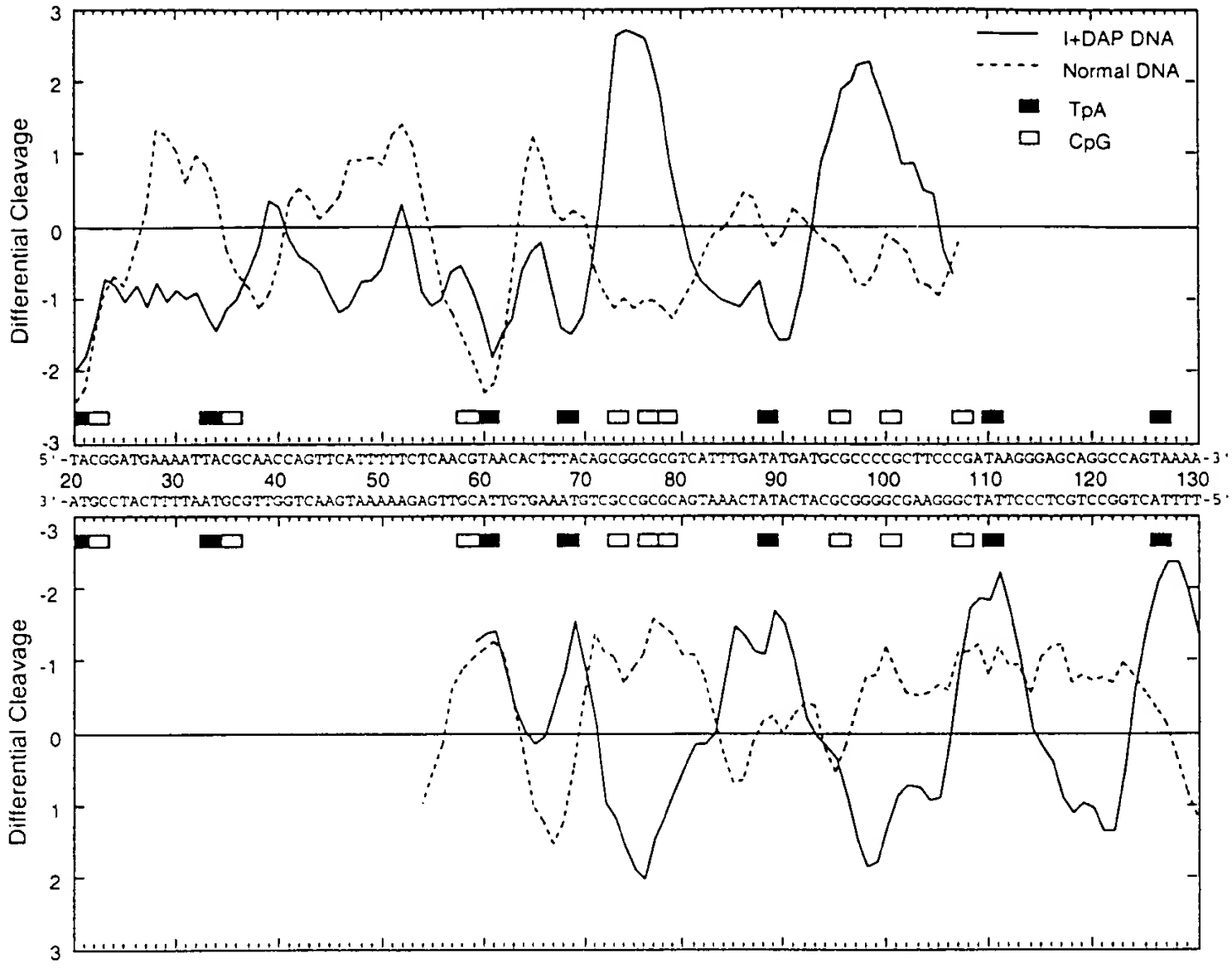


Figure 5. Differential cleavage plots comparing the susceptibility of *tyr*T(A93) DNA to cutting by DNAase I in the presence of 10 μM echinomycin. The dashed and continuous curves refer to natural and inosine plus DAP-substituted DNA respectively, as for Figure 4. The upper panel shows differential cleavage of the Watson strand, the lower panel that of the complementary Crick strand. The ordinate scales for the two strands are inverted, so that deviation of points towards the lettered sequence (negative values) corresponds to a ligand-protected site and deviation away (positive values) represents enhanced cleavage. The filled rectangles near the indicated dinucleotide steps show the positions of the TpA steps and the open rectangles show the positions of the CpG steps. Other details as for Figure 4.

position 82 can any departure from strict adherence to this rule be discerned: at very low echinomycin concentrations (≤1 μM) the band corresponding to cutting at this step is clearly enhanced but by 10–20 μM the enhancement is lost, suggestive of incipient protection (footprinting) at this site. This looks like a classic instance of secondary binding, characterised by diminished specificity, to a site for which the antibiotic has lower affinity (and which lacks one critical 2-amino recognition element) such as is commonly seen with many DNA-binding drugs (6).

A full analysis of the behaviour of actinomycin D has yielded equivalent results (31): the binding sites in doubly inosine plus DAP-substituted DNA lie squarely over the DAPpT steps and all such steps form part of a binding site for the antibiotic, whereas the GpC step is the essential component of canonical recognition sequences in natural DNA. Not so for nogalamycin, however: the sequences to which this antibiotic binds in the two types of DNA are not greatly different though the intensity of the footprinting profile is markedly enhanced at specific sites particularly on the lower (Crick) strand (Fig. 7). A previous exhaustive footprinting study with several DNA restriction fragments (28) concluded that most nogalamycin binding sites are located near regions of alternating purine–pyrimidine sequence, most commonly associated with the dinucleotide steps TpG (CpA) and GpT (ApC), suggesting that the preferred antibiotic binding sites may contain all four nucleotides and/or that peculiarities of the dynamics of DNA conformation at alternating sequences may be critical for nogalamycin binding. Given its unusual mode of 'threading' intercalation which probably involves local disruption of base-pairing and results in conspicuously slow association and dissociation kinetics (32,33) the results in Figure 7 are perhaps unsurprising. It is likely that the binding of this ligand is dominated by the structural and dynamic features of DNA which may well be broadly similar for natural and doubly-substituted DNA molecules. For example, the alternation of purines and pyrimidines at particular sites remains the same in both types of DNA. At all events, there is rather poor correspondence between the sites protected from nuclease cleavage by nogalamycin and the TpG (CpA) and GpT (ApC) steps in natural DNA or TpI (CpDAP) and IpT (DAPpC) steps in substituted DNA, so the process of site recognition by nogalamycin appears to proceed to a large extent independently of the placement of the purine 2-amino group. It is tempting to speculate that the preferred nogalamycin binding sites are determined chiefly by substituents lying in the major groove of the DNA helix, which is where one of the bulky sugar substituents of the antibiotic must come to lie (27), and consequently the best sites are to be found at the same places in both types of DNA in Figure 7 although the ease of binding to such sites reflected in the binding kinetics may be

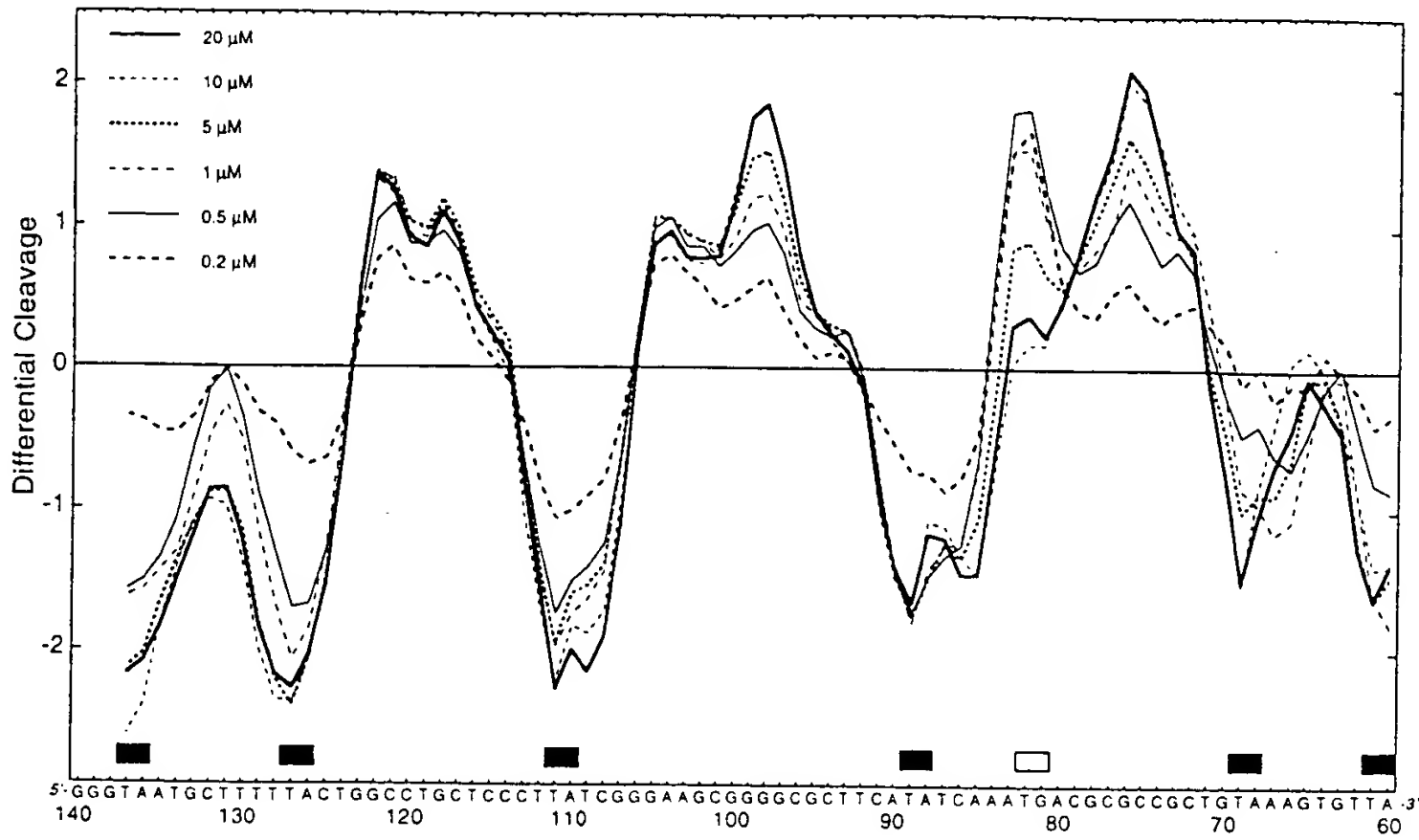


Figure 6. Differential cleavage plots showing the concentration-dependence of footprinting by 0.2–20 μM echinomycin on the Crick (sense) strand of *tyr*T(A93) DNA containing inosine plus DAP residues. Positive and negative values represent, respectively, enhanced or diminished DNAase I cutting efficiency at each internucleotide bond. Note that the sequence shown corresponds to that of natural DNA, though here the A residues are replaced by DAP and the G residues by inosine. Black boxes show the locations of the TpDAP steps. A white box is drawn over the TpI step at position 82.

much affected, witness the exaggerated excursions of the differential cleavage plot for the substituted DNA.

The findings summarised in Figures 3 and 4 establish that the purine 2-amino group characteristic of guanine nucleotides serves as both a positive and negative effector to determine ligand binding sites, and thereby dominates the sequence-recognition process. This interpretation is amplified by parallel observations on the interaction of the same antibiotics with singly inosine (11) or DAP-substituted (34) DNA molecules. The positive role of the 2-amino group as a necessary component of binding sites for GC-selective antibiotics is confirmed by the failure of echinomycin, actinomycin, chromomycin and mithramycin to footprint on inosine-containing DNA, plus their redistribution on to new sites in DAP-substituted DNA. Conversely, the negative signal given by the 2-amino group as a marker of sites unavailable for binding AT-selective antibiotics is rendered obvious by the failure of netropsin and distamycin to footprint on DAP-containing DNA, whereas they seem to bind all over the inosine-substituted polymer.

Quantitative analysis of ligand-site interactions backs up these conclusions. The full concentration-dependence profile (35) for each antibiotic has been determined at all binding sites and a few examples are illustrated in Figure 8. With echinomycin, the footprinting on natural DNA at the canonical CpG steps and the corresponding enhancement of DNAase I cleavage at TpA steps occur with half-maximal effect at concentrations (C_{50}) ~2–5 μM. By contrast, the footprinting at TpDAP steps in inosine plus DAP-substituted DNA, and enhancement of cutting at CpI steps, take place at much lower concentrations: C_{50} = 1 μM or below. The same is true for DNA substituted only with DAP. We must conclude that the new DAP-containing binding sites are superior to the canonical CpG-containing sites. Under the conditions of these footprinting experiments a large fraction of the added ligand is likely to remain free, such that C_{50} values will approximate to thermodynamic dissociation constants for binding to individual sites (35). On this basis we estimate that the binding constant for echinomycin at several DAP-containing sites must be enhanced by a factor of ten or more. With netropsin and distamycin no such potentiation of effect occurs: C_{50} values for footprinting at the newly created IC-containing binding sites, and for enhanced nuclease cleavage at DAP•T-rich clusters, are generally somewhat higher than those required to produce equivalent effects at the canonical sites in normal DNA (Fig. 8). With actinomycin and the other antibiotics the apparent affinity for new binding sites in doubly-substituted DNA is generally little different from that measured for the natural *tyr*T fragment.

DISCUSSION

What is the mechanism by which the purine 2-amino group dictates where sequence-selective small molecules bind to DNA? It could be mediated via direct contact between the interacting species involving the formation of hydrogen bonds together with elements of steric complementarity—the so-called 'digital' readout (36). Alternatively it could occur by a kind of 'analogue' readout in which the favoured binding sites are recognised by virtue of some conformational property such as groove width which is only indirectly influenced by the location of the critical 2-amino group. We have evidence from comparing the reactivity of inosine and DAP-substituted DNAs towards structure-sensitive probes that groove width is a relevant parameter which can be strongly affected by re- positioning the purine 2-amino group (37). Deformability of the helix is another property which is likely to be affected. Either or both of these features of DNA helical structure could conspire to determine what nucleotide sequences,

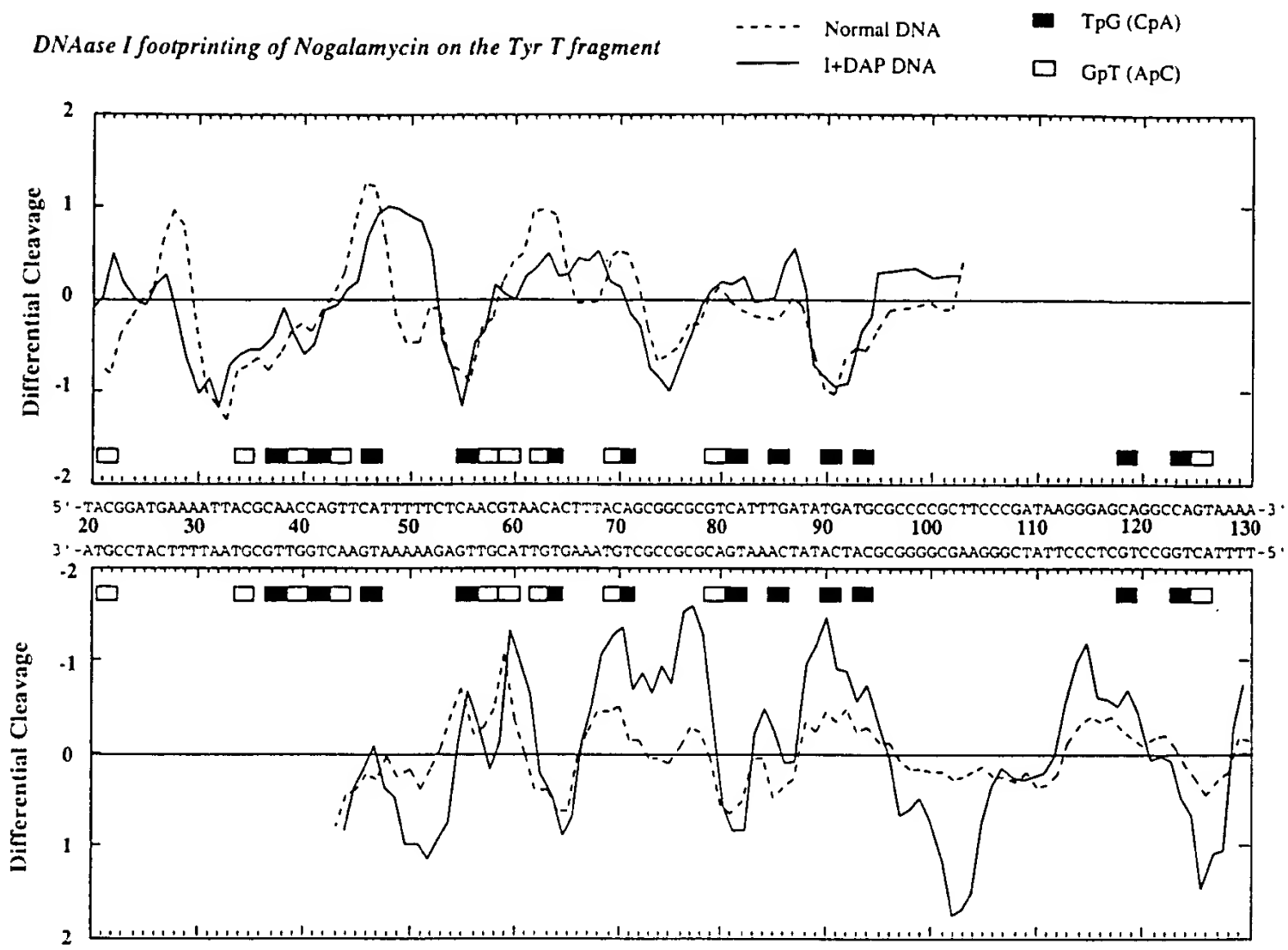


Figure 7. Differential cleavage plots comparing the susceptibility of *tyr*T(A93) DNA to cutting by DNAase I in the presence of 5 μM nogalamycin. Details as for Figures 4 and 5. Black rectangles indicate the TpG (CpA) steps, open rectangles the GpT (ApC) steps.

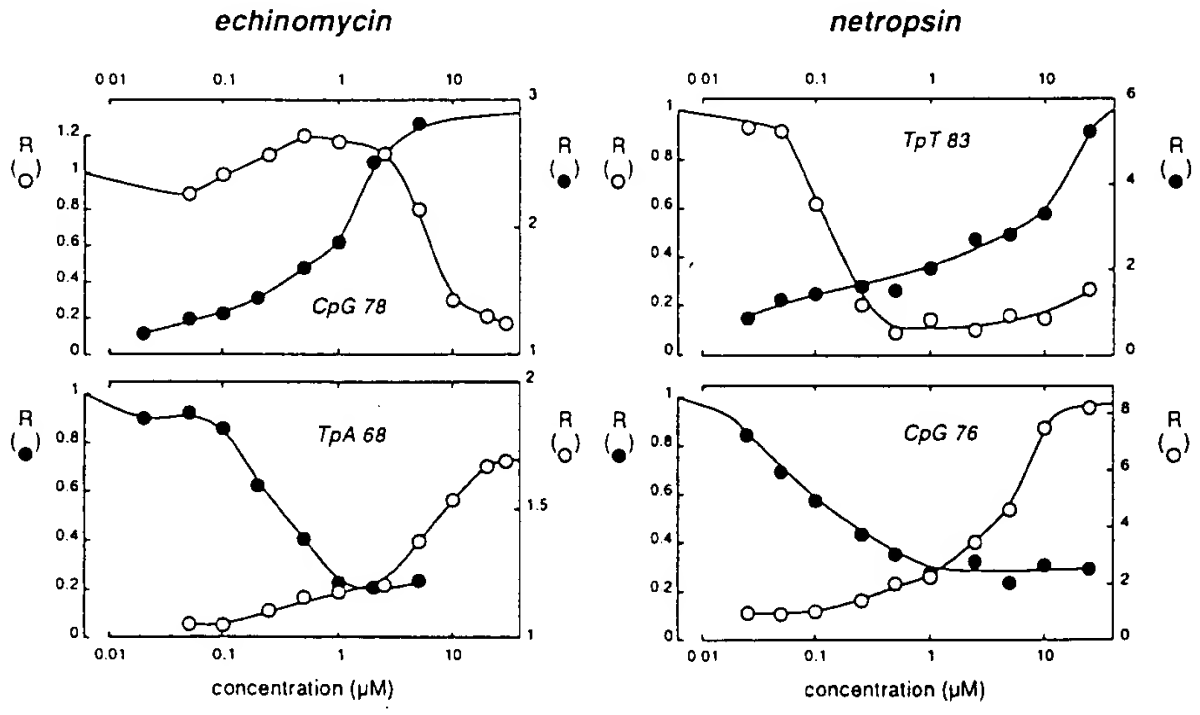


Figure 8. Footprinting plots (35) for selected bonds in the normal *tyr*T(A93) fragment (open circles) and its inosine plus DAP-substituted homologue (filled circles). The relative band intensity R corresponds to the ratio I_c/I_0 where I_c is the intensity of the band at a ligand concentration c and I_0 is the intensity of the same band in the absence of antibiotic. Ligand concentrations are plotted on a logarithmic abscissa and each plot is linked to the appropriate scale on the left or right ordinate.

in a particular context, might constitute an acceptable ligand binding site. Indeed, for a given ligand the part played by any single factor ultimately under the control of the purine 2-amino group could vary at different potential binding sites. At all events, based on our present findings we are confident that efforts to design DNA sequence-specific ligands using various motifs, all crucially dependent upon exploiting principles of molecular recognition, can be placed on a firmer footing.

ACKNOWLEDGEMENTS

We thank Dean Gentle for his invaluable technical assistance. This work was supported by grants (to MJW) from the Cancer Research Campaign, the Wellcome Trust and the Association for International Cancer Research. CB thanks the INSERM and the Sir Halley Stewart Trust.

REFERENCES

1 Dervan, P.B. (1986) *Science* **232**, 464–471.
2 Hurley, L.H. (1989) *J. Med. Chem.* **32**, 2027–2033.
3 Neidle, S. and Waring, M. J. *Molecular Aspects of Anticancer Drug–DNA Interactions*, Vols 1 (1993) and 2 (1994), Macmillan, London.
4 Waring, M.J. (1968) *Nature* **219**, 1320–1325.
5 Larsen, T.A., Kopka, M.L. and Dickerson, R.E. (1991) *Biochemistry* **30**, 4443–4449.
6 Gale, E.F., Cundliffe, E., Reynolds, P.E., Richmond, M.H. and Waring, M.J. (1981) *The Molecular Basis of Antibiotic Action*, 2nd ed., pp. 258–401, Wiley, London.
7 Wang, A.H.J. (1993) in reference (3), Vol. 1, pp 32–53.
8 Low, C.M.L., Drew, H.R. and Waring, M.J. (1984) *Nucleic Acids Res.* **12**, 4865–4877.
9 Drew, H.R., Weeks, J.R. and Travers, A.A. (1985) *EMBO J.* **4**, 1025–1032.
10 Sayers, E.W. and Waring, M.J. (1993) *Biochemistry* **32**, 9094–9107.
11 Marchand, C., Bailly, C., McLean, M.J., Moroney, S.E. and Waring, M.J. (1992) *Nucleic Acids Res.* **20**, 5601–5606.
12 Bloch,W. (1991) *Biochemistry* **30**, 2735–2747.
13 Maniatis,T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
14 Johnston, R.F., Pickett, S.C. and Barker, D.L. (1990) *Electrophoresis* **11**, 355–360.
15 Waring, M.J. and Wakelin, L.P.G. (1974) *Nature* **252**, 653–657.
16 Goodisman, J., Rehfuss, R., Ward, B. and Dabrowiak, J.C. (1992) *Biochemistry* **31**, 1046–1058.
17 Kamitori, S. and Takusagawa, F. (1992) *J. Mol. Biol.* **225**, 445–456.
18 Ward, B., Rehfuss, R., Goodisman, J. and Dabrowiak, J.C. (1988) *Biochemistry* **27**, 1198–1205.
19 Van Dyke, M.W., Hertzberg, R.P. and Dervan, P.B. (1983) *Proc. Natl. Acad. Sci. USA* **79**, 5470–5474.
20 Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R.E. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1376–1380.
21 Van Dyke, M.W. and Dervan, P.B. (1983) *Biochemistry* **22**, 2373–2377.
22 Gao, X. and Patel, D.J. (1989) *Biochemistry* **29**, 10940–10956.
23 Sastry, M. and Patel, D.J. (1993) *Biochemistry* **32**, 6588–6604.
24 Keniry, M.A., Banville, D.L., Simmonds, P.M. and Shafer, R. (1993) *J. Mol. Biol.* **231**, 753–767.
25 Egli, M., Williams, L.D., Frederick, C.A. and Rich, A. (1991) *Biochemistry* **30**, 1364–1372.
26 van Houte, L.P.A., van Garderen, C.J. and Patel, D.J. (1993) *Biochemistry* **32**, 1667–1674.
27 Smith, C.K., Davies, G.J., Dodson, E.J. and Moore, M.H. (1995) *Biochemistry* **34**, 415–425.
28 Fox, K.R. and Waring, M.J. (1986) *Biochemistry* **25**, 4349–4356.
29 Portugal, J. and Waring, M.J. (1987) *Eur. J. Biochem.* **167**, 281–289.
30 Portugal, J. and Waring, M.J. (1988) *Biochim. Biophys. Acta* **949**, 158–168.
31 Waring, M.J. and Bailly, C. (1994) *Gene* **149**, 69–79.
32 Fox, K.R. and Waring, M.J. (1984) *Biochim. Biophys. Acta* **802**, 162–168.
33 Fox, K.R., Brassett, C. and Waring, M.J. (1985) *Biochim. Biophys. Acta* **840**, 383–392.
34 Bailly, C., Marchand, C. and Waring, M.J. (1993) *J. Am. Chem. Soc.* **115**, 3784–3785.
35 Dabrowiak J.C. and Goodisman J. (1989) in *Chemistry and Physics of DNA–Ligand Interactions* (ed. Kallenbach, N. R.) pp. 143–174, Adenine Press.
36 Nielsen, P.E. (1991) *Bioconjugate Chem.* **2**, 1–12.
37 Bailly, C., Mollegaard, N.E., Nielsen, P.E. and Waring, M.J. (1995) *EMBO J.*, in press.

Identifying RNA Minor Groove Tertiary Contacts by Nucleotide Analogue Interference Mapping with N^2-Methylguanosine†

Lori Ortoleva-Donnelly, Matthew Kronman, and Scott A. Strobel*

Department of Molecular Biophysics and Biochemistry, Yale University, 260 Whitney Avenue, New Haven, Connecticut 06520

Received March 30, 1998; Revised Manuscript Received July 2, 1998

ABSTRACT: Nucleotide analogue interference mapping (NAIM) is a general biochemical method that rapidly identifies the chemical groups important for RNA function. In principle, NAIM can be extended to any nucleotide that can be incorporated into an in vitro transcript by an RNA polymerase. Here we report the synthesis of 5′-*O*-(1-thio)-N^2-methylguanosine triphosphate ($m^2G\alpha S$) and its incorporation into two reverse splicing forms of the *Tetrahymena* group I intron using a mutant form of T7 RNA polymerase. This analogue replaces one proton of the N2 exocyclic amine with a methyl group, but is as stable as guanosine (G) for secondary structure formation. We have identified three sites of $m^2G\alpha S$ interference within the *Tetrahymena* intron: G22, G212, and G303. All three of these guanosine residues are known to utilize their exocyclic amino groups to participate in tertiary hydrogen bonds within the ribozyme structure. Unlike the interference pattern with the phosphorothioate of inosine (IαS, an analogue that deletes the N2 amine of G), $m^2G\alpha S$ substitution did not cause interference at positions attributable to secondary structural stability effects. Given that the RNA minor groove is likely to be widely used for helix packing, $m^2G\alpha S$ provides an especially valuable reagent to identify RNA minor groove tertiary contacts in less well-characterized RNAs.

The wide and shallow minor groove of the RNA A-form double helix appears to be commonly used for helix packing interactions that are necessary for the formation of RNA tertiary structure (*1*). Structural motifs involving the RNA minor groove include the "ribose zipper" and the "wobble receptor", both of which employ nucleotide functional groups unique to the minor groove surface (*2*, *3*). Unfortunately, the minor groove is difficult to analyze biochemically because the chemical reagents typically used in footprinting and interference studies are not informative for the minor groove functional groups. One reagent that is reactive with these groups is kethoxal, which forms a covalent bridge between the N2 exocyclic amine and N1 groups of G when both functional groups are accessible (*4*, *5*). While this reagent is valuable for identifying unpaired Gs within a sequence, kethoxal cannot be used to identify tertiary interactions within helical segments of the RNA. Thus, despite the importance of the minor groove in RNA helix packing, there are significant deficiencies in our ability to explore the minor groove face of the helix with the probing reagents that are currently available.

Nucleotide analogue interference mapping (NAIM)[1] is an efficient method to define the chemical basis of RNA function (*6*−*9*). In principle, the method is generalizable to a wide variety of nucleotide derivatives. In this approach, the nucleotide analogue is chemically tagged with a phosphorothioate linkage and randomly incorporated into the RNA transcript. The sites of phosphorothioate incorporation are detected by cleavage of the linkage with iodine (*10*). This makes it possible to simultaneously, yet individually, quantitate every position where the nucleotide was incorporated into the RNA with an interference assay that is as simple as RNA sequencing. This approach has been used with 2′-deoxy and 2′-methoxy analogues to identify the 2′-OH groups in tRNA and RNase P that are essential for binding (*6*, *7*, *11*). It has also been employed with inosine and a series of eight adenosine derivatives including 7-deazaadenosine, purine riboside, diaminopurine riboside, and N^6-methyladenosine to explore group I intron catalysis (*3*, *8*, *9*).

The *Tetrahymena* ribozyme provides an ideal system to further develop NAIM methodology (Figure 1). The intron catalyzes two consecutive transesterification reactions in the course of RNA self-splicing (*12*). It can also catalyze the reverse of these two reactions, resulting in ligation of the exon back onto the intron (Figure 2A,B) (*13*−*16*). By using a radiolabeled oligonucleotide analogue of the exons, the active ribozymes in the population become site-specifically-labeled upon exon ligation, which provides a remarkably simple assay to probe for sites of interference (Figure 2C) (*8*). Furthermore, the *Tetrahymena* ribozyme ranks as possibly the best characterized large RNA, because a crystal structure of the P4−P6 domain of the intron (160 of the 414

† This work was supported by NSF Grant CHE-9701787, a Beckman Young Investigator Award, a Searle Scholar Award, and a Junior Faculty Research Award from the American Cancer Society.

* To whom correspondence should be addressed. Phone: (203)-432-9772. FAX: (203)432-5175. Email address: strobel@csb.yale.edu.

[1] Abbreviations: m^2G, N^2-methylguanosine; $m^2G\alpha S$, 5′-*O*-(1-thio)-N^2-methylguanosine monophosphate; $m^2GTP\alpha S$, 5′-*O*-(1-thio)-N^2-methylguanosine triphosphate; GαS, 5′-*O*-(1-thio)guanosine; GTPαS, 5′-*O*-(1-thio)guanosine triphosphate; IαS, 5′-*O*-(1-thio)inosine monophosphate; NAIM, nucleotide analogue interference mapping; dT, thymidine; rT, 5-methyluridine; dT(-1)S, CCCUC(dT)AAAAA; dT(-1)P, CCCUC(dT); rT(-1)P, CCCUC(rT); SDS, sodium dodecyl sulfate; IGS, internal guide sequence.

S0006-2960(98)00723-5 CCC: $15.00
Published on Web 08/28/1998

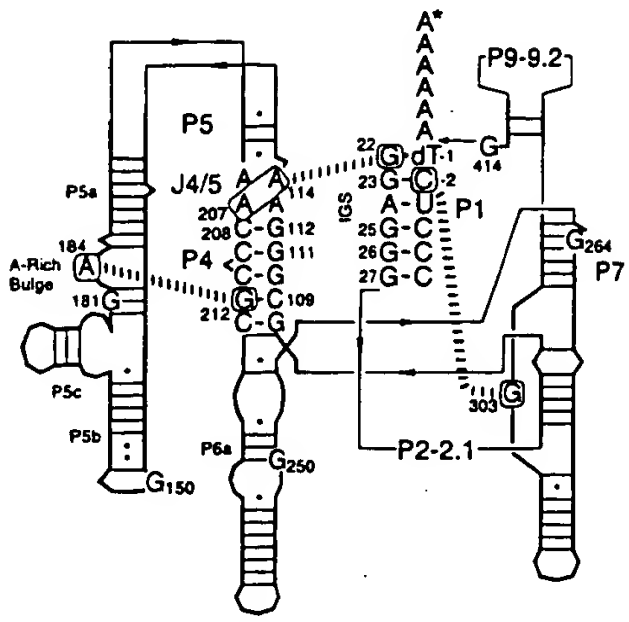


FIGURE 1: Secondary structure of the L-21 G414 form of the *Tetrahymena* group I intron. Nucleotides discussed in the text are shown, as are the names of the helical and single-stranded regions of the RNA. Other nucleotides are depicted as heavy lines. Connectivity within the ribozyme sequence is shown as thin lines, and the tertiary hydrogen bonds formed by the three Gs that display $m^2G\alpha S$ interference are shown as dashed lines. This ribozyme binds the oligonucleotide dT(-1)S, CCCUC(dT)AAAAA, and transfers the AAAAA onto G414 at the 3′-end of the intron in a reaction analogous to the reverse of the second step of splicing (*15, 16*).

nucleotides) is available, and three-dimensional models of the rest of the intron have been proposed based upon phylogenetic and other biochemical experiments (*2, 17*−*20*). Thus, interference in this ribozyme system provides a basis set to calibrate NAIM methodology prior to its application on less well-characterized RNAs. Yet, even within this RNA, the interference results can help to refine and improve our understanding of group I intron structure and catalysis.

Previous mapping experiments of the *Tetrahymena* group I intron with the phosphorothioate of inosine (IαS), a G analogue that replaces the N2 exocyclic amine with a proton (Figure 3), revealed several sites that interfered with ribozyme activity (*8*). Interference at many of these sites is unlikely to indicate direct participation of the amine in tertiary hydrogen bonding, but rather reflects a loss of duplex stability. In hopes of developing a better analogue to analyze the role of the minor groove exocyclic amine in tertiary structure formation, we synthesized the 5′-*O*-(1-thio)-N^2-methylguanosine triphosphate ($m^2GTP\alpha S$) and utilized it in NAIM. Instead of deleting the amine, this analogue replaces one of the amino protons with a methyl group (Figure 3). We find that the sites of interference throughout the *Tetrahymena* group I intron are exclusively at positions where the exocyclic amine of G is known to participate in long-range tertiary hydrogen bonds. Thus, $m^2G\alpha S$ provides a means to identify tertiary interactions within the RNA minor groove, a region that is relatively uninformative using previously available biochemical methodologies.

METHODS

Synthesis of $m^2GTP\alpha S$. O^6-[2-(Nitrophenyl)ethyl]-N^2-methylguanosine was prepared as previously described (*21*−*23*). N^2-Methylguanosine was synthesized by treating O^6-[2-(nitrophenyl)ethyl]-N^2-methylguanosine (1.0 g, 2.3 mmol) with 1,8-diazabicyclo[5.4.0]undec-7-ene (5 mL) at room temperature overnight. Water (50 mL) was added to the reaction, and the aqueous phase was extracted with CH_2Cl_2

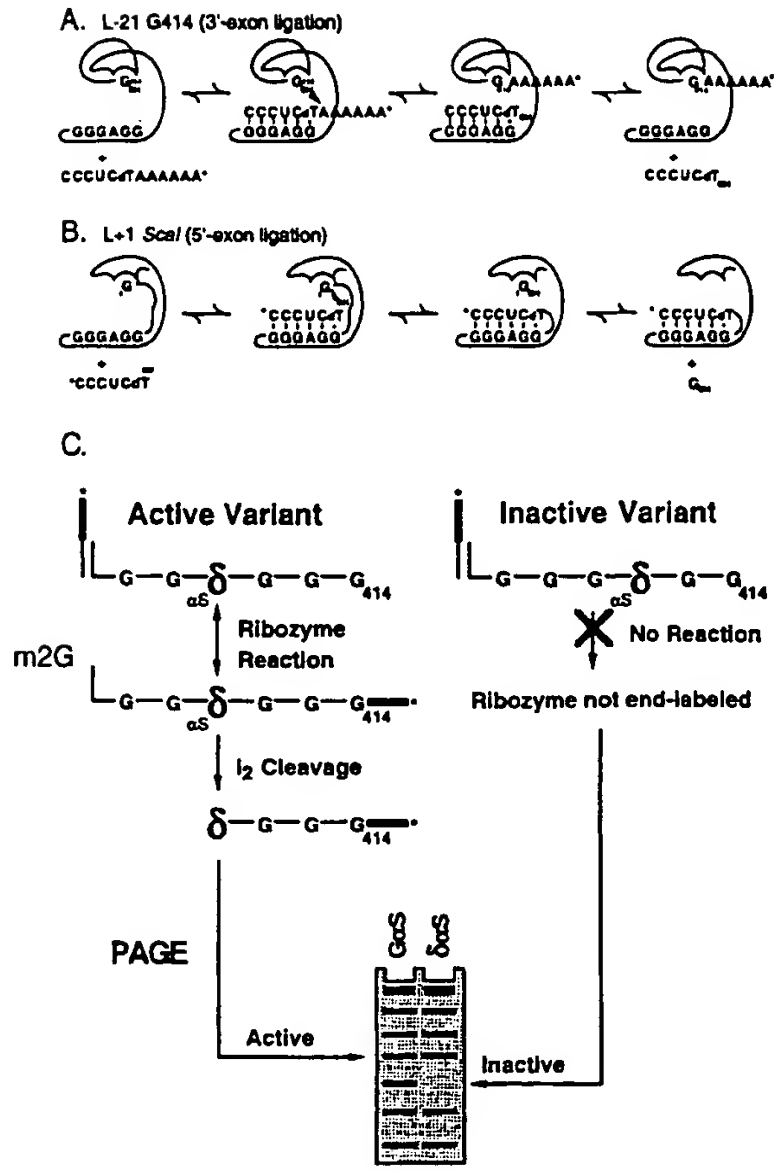


FIGURE 2: (A) Scheme for the reaction of the L-21 G414 ribozyme with oligonucleotide substrate. This reaction is analogous to the reverse of the second step of splicing (*13, 15, 16*). The ribozyme binds the substrate to form the P1 helix, which docks into the active site. The terminal guanosine (G414) nucleophilically attacks the substrate and transfers the 3′-end of the oligonucleotide onto the 3′-end of the intron. The equilibrium constant for the chemical step of this reversible reaction is approximately 1 (*16*). This reaction selectively 3′-end-labels the active ribozymes in the population when a 3′-end-labeled substrate (∗) is used. (B) Scheme for L+1 *ScaI* ribozyme reaction analogous to the reverse of the first step of splicing (*14*). The ribozyme binds the 5′-exon oligonucleotide analogue, where the 3′-OH of the exon attacks the 3′-phosphate of G1, releasing G1 as a free nucleotide and adding the 5′-exon onto the ribozyme. This reaction selectively 5′-end-labels the active ribozymes in the population when a 5′-end-labeled (∗) oligonucleotide is used. (C) Scheme for the identification of the chemical groups important for RNA activity by NAIM (*8*). The phosphorothioate-tagged nucleotide analogue (indicated as $\delta\alpha S$) is randomly incorporated into the transcript in place of G. If $m^2G\alpha S$ does not interfere with function at a particular position (left side), then ribozymes with the analogue at that site perform the ligation reaction and become radiolabeled. If $m^2G\alpha S$ disrupts activity (right side), then the subset of ribozymes that have $m^2G\alpha S$ incorporated at the susceptible site do not perform the ligation reaction and are not radiolabeled. Cleavage of the phosphorothioate linkages by treatment with iodine and resolution of the cleavage products by PAGE produce a sequencing ladder with gaps that correspond to sites intolerant of $m^2G\alpha S$ substitution. $G\alpha S$ serves as a control to ensure that loss of activity is not due to the phosphorothioate group. Unreacted RNA is also 5′-end-labeled to ensure that the gap in the sequencing ladder is not due to lack of $m^2G\alpha S$ incorporation at a given site (not shown).

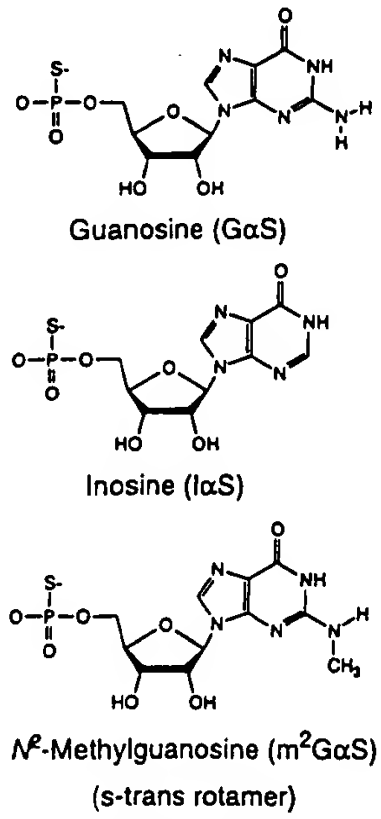


FIGURE 3: Parental G nucleotide and the two nucleotide analogues used for NAIM in this work. Each nucleotide is shown as the monophosphate derivative, the form in which it is incorporated during transcription. N^2-Methylguanosine can adopt either the s-trans (shown) or the s-cis rotamer.

(3 × 100 mL) and with ether (3 × 100 mL). The water was removed by rotary evaporation and the reaction dried overnight in vacuo. Authentic N^2-methylguanosine (*24*) was obtained as an off-white solid (340 mg, 50% yield) by recrystallization of the residue with hot methanol.

Synthesis of the 5′-*O*-(1-thio)-N^2-methylguanosine triphosphate followed the general procedure outlined by Arabshahi and Frey (*25*). N^2-Methylguanosine (50 mg, 0.17 mmol) was dried under vacuum at 110 °C for 16 h and dissolved in triethyl phosphate (2 mL). Trioctylamine (83 μL, 0.19 mmol, 1.1 equiv) and thiophosphoryl chloride (20 μL, 0.19 mmol, 1.1 equiv) were added to the reaction and stirred under argon at room temperature for 30 min to form 5′-*O*-(1-thio-1,1-dichloro)phosphoryl-N^2-methylguanosine. The reaction was about 40% complete based upon thin-layer chromatography (TLC) on cellulose plates using 0.5 M LiCl (aq) as the solvent system. This was converted directly to the triphosphate by addition of tributylammonium pyrophosphate (105 mg, 0.34 mmol, 2.0 equiv) in triethyl phosphate (3 mL) and stirring at room temperature for an additional 30 min. Formation of the N^2-methylguanosine 5′-*O*-(1-thio)-*cyclo*-triphosphate was monitored by silica TLC using 6:3:1 1-propanol, ammonium hydroxide, water as the solvent system. In this system, the triphosphate had an R_f of 0.2 compared to 0.6 for the monophosphate and 0.8 for the free nucleoside. The triphosphate was precipitated by addition of excess triethylamine (2.5 mL), centrifuged, and decanted, and the residue was dissolved in aqueous triethylammonium bicarbonate (TEAB) (50 mM, pH 7.5, 10 mL). The crude product was left at room temperature overnight to achieve hydrolytic ring opening of the cyclic triphosphate. Purification by DEAE-A25 Sephadex chromatography using a linear gradient of 0.05−0.8 M TEAB afforded 5′-*O*-(1-thio)-N^2-methylguanosine triphosphate (m^2GTPαS) as a diastereomeric mixture in 22% yield. The triphosphate eluted at approximately 0.6 M TEAB. ^{31}P NMR (H_2O): 43.37 (m), −6.87 (d), −23.00 (t); λ_{max} 254 nm; ϵ_{254} 13 000.

Transcriptional Incorporation of m^2GαS. Plasmid templates for RNA transcription were prepared by ionic exchange chromatography (Qiagen), digested with the appropriate restriction enzyme, phenol extracted, and ethanol precipitated. pUCL-21G414 was cut with *Ear*I, and pUCL+1 was cut with *Sca*I (*8*). m^2GαS was randomly incorporated into the L-21 G414 or L+1 *Sca*I forms of the intron by in vitro transcription using the wild-type or Y639F mutant form of T7 RNA polymerase (*26*). RNAs were transcribed in 40 mM Tris-HCl, pH 7.5, 4 mM spermidine, 10 mM DTT, 15 mM $MgCl_2$, 0.05% Triton X-100, 0.05 μg/μL DNA template, and 1 mM CTP, UTP, and ATP. Using the pUCL-21G414 plasmid, various concentration ratios of m^2GTPαS (0.1, 0.5, 1.0, and 2.0 mM) and GTP (0.1, 0.5, and 1.0 mM) were tested to identify a transcription condition that gave approximately 5% m^2GαS incorporation as defined by comparison to a transcript made with 50 μM GTPαS (S_P diastereomer only) and 1 mM GTP (*27*). Following this determination, the L+1 *Sca*I RNA was transcribed using 1.0 mM m^2GTPαS, 0.5 mM GTP, and the Y639F polymerase. All the RNAs were purified by PAGE (6% denaturing), eluted into 10 mM Tris, pH 7.5, 0.1 mM EDTA (TE) overnight at 4 °C, precipitated with NaCl and ethanol (−80 °C for 2 h), resuspended in TE, and stored at −20 °C.

5′- and 3′-End-Labeling of RNA. Three oligonucleotides were utilized in the interference mapping experiments. dT(-1)S, CCCUC(dT)AAAAA (20 pmol), was radiolabeled at the 3′-end with [α-^{32}P]cordycepin by yeast poly(A) polymerase (Amersham) (*28*) and used as a substrate for the 3′-exon ligation experiments. dT(-1)P, CCCUC(dT), and rT(-1)P, CCCUC(rT), were each radiolabeled at the 5′-end with [γ-^{32}P]ATP by T4 polynucleotide kinase and used as substrates for the 5′-exon ligation experiments. Both RNAs were purified by PAGE (10% nondenaturing), eluted in TE, and used in the ligation assays without further treatment.

L-21 G414 RNAs (2.5 pmol) were treated with calf intestinal alkaline phosphatase (2 units, 30 min, 37 °C) to remove the 5′-phosphate and heated to 85 °C for 15 min to inactivate the phosphatase. The RNAs were 5′-end-labeled using T4 polynucleotide kinase (5 units) and [γ-^{32}P]ATP at 37 °C for 30 min. The radiolabeled RNAs were purified by PAGE (6% denaturing) and eluted into 0.1% SDS in TE overnight. The SDS was removed by extraction with 1 volume of phenol/chloroform (1:1), and the RNAs were precipitated with NaCl and ethanol, centrifuged, decanted, and resuspended in 50 μL of TE. The specific activities of the 5′-end-labeled RNAs (cpm/μL) were normalized by scintillation counting. The RNAs were cleaved by the addition of 0.1 volume of 100 mM iodine in ethanol and heated to 90 °C for 1 min, and the cleavage products were resolved by PAGE on a 6% or a 5% denaturing gel. Several loadings of the same reaction samples were electrophoresed for variable amounts of time (from 1 to 5 h) to maximally resolve each region of the sequence.

Despite several attempts, we were unable to obtain clean sequence information by 5′-end-labeling the L+1 *Sca*I RNA due to some heterogeneity at the 5′-end of the L+1 *Sca*I RNA. To obtain information for the positions near the 5′-end of the transcript that is comparable to that derived from the 5′-end-labeled control, the L+1 *Sca*I RNA was incubated

with a highly reactive substrate, rT(-1)P, under permissive reaction conditions (10 mM $MgCl_2$, 50 °C, 50 mM HEPES, pH 7.0) for 10 min. The faster reacting ribose substrate reduces the amount of interference observed throughout the length of the RNA and provides a minimum estimate as to the extent of analogue incorporation at each site. This version of the "5′-end-labeled control" was used in the L+1 *Sca*I interference calculations for positions in the internal guide sequence (IGS) and P2 helix (*8*).

m^2GαS Interference Mapping of the Tetrahymena Ribozyme. Interference mapping of m^2GαS was performed using 50 nM ribozyme and 10 nM oligonucleotide substrate in a buffer containing 4 mM $MgCl_2$, 50 mM HEPES, pH 7.0, at 50 °C for 10 min. The L-21 G414 ribozyme was incubated with 3′-end-labeled dT(-1)S, and the L+1 *Sca*I ribozyme was incubated with 5′-end-labeled dT(-1)P. The ligation reactions were stopped by the addition of 1 volume of urea loading buffer (8 M urea, 50 mM EDTA, 0.01% bromophenol blue, 0.01% xylene cyanol) and worked up with iodine as described above. The reaction containing no iodine was run in parallel to confirm that the cleavage pattern was specific to the iodine treatment and not due to nonspecific degradation. The transcriptional efficiencies of m^2GαS incorporation were determined using the 3′-exon ligation reaction of L-21 G414 RNAs with dT(-1)S. Relative efficiencies were calculated by comparing the intensity of the cleavage products throughout the length of the intron to those of the GαS control (5% incorporation standard).

Interference Quantitation. Peak intensities for both the parental nucleotide (GαS) and the nucleotide analogue (m^2GαS) were quantitated by PhosphorImager analysis at each position for the 3′-exon ligation or 5′-exon ligation experiments and the 5′-end-labeled control. The extent of interference at each position was calculated by substituting the band intensities at each nucleotide position into the equation:

$$\text{Interference} = \frac{\text{GαS ligation reaction}/m^2\text{GαS ligation reaction}}{\text{GαS labeled control}/m^2\text{GαS labeled control}} \quad (1)$$

The resulting interference value normalizes for phosphorothioate effects assumed to be equivalent for both GαS and m^2GαS. It also controls for variability in the extent of analogue incorporation or reactivity with iodine at each position. All the interference values were further normalized to account for differences in loading and extent of reaction between lanes by calculating the average interference value at all positions in the RNA that were within two standard deviations from the mean and dividing each individual interference value by the normalized average (the averages ranged from 0.8 to 1.2). This resulted in an interference κ value for each position in both the 3′-exon and 5′-exon ligation reactions. A κ value of 1 indicates that there is no effect of substituting the analogue at that site, a value greater than 1 indicates inhibition of activity, and a value less than 1 indicates that activity is enhanced by analogue substitution at that site.

RESULTS

Transcriptional Incorporation of m^2GαS. The α-phosphorothioate-tagged triphosphate of m^2G was synthesized for use in NAIM. To be useful in this method, it must be efficiently and accurately incorporated into an RNA by in vitro transcription. Initial efforts to incorporate m^2GαS were unsuccessful using T7 RNA polymerase, probably because the methyl group occupies a prominent position in the minor groove that is likely to be involved in the error-reading mechanism of the polymerase (*29*). No incorporation was observed even at high ratios of m^2GTPαS to GTP (data not shown). We had previously obtained and overexpressed a Y639F point mutant of T7 RNA polymerase to perform NAIM with 2′-deoxynucleotides. This mutation was reported to cause reduced selectivity for the 2′-position of the nucleotide during transcription (*26*). The mutant polymerase efficiently incorporates 2′-deoxy, 2′-methoxy, 2′-fluoro, and 2′-thio nucleotide triphosphates into RNA transcripts (*9*, *30*–*32*). It seemed reasonable to expect that this polymerase might also have enhanced tolerance for an additional methyl group in the helical minor groove.

We repeated the transcriptions using the Y639F polymerase and found that m^2GαS was efficiently incorporated into the L-21 G414 RNA. An incorporation level of approximately 5% was achieved using a ratio of 1.5 mM m^2GTPαS to 0.5 mM GTP. Based upon iodine treatment of the 5′-end-labeled L-21 G414 transcript, m^2GαS was incorporated at every G position within the RNA, and no significant incorporation was detected at any non-G sites within the sequence (Figure 4A, lanes 2 and 3). Furthermore, the efficiency of incorporation at each G was generally equivalent to that seen for the GαS standard.

m^2GαS Interference Mapping. The sites of m^2GαS incorporation that interfere with intron activity were mapped using the L-21 G414 ribozyme (*8*). pUCL-21G414 encodes a form of the group I intron that includes the terminal G414, but lacks the first 21 nucleotides of the intron. In the presence of an oligonucleotide substrate analogous to the 5′-3′ ligated exon product, dT(-1)S, this intron can perform the reverse of the second step of splicing by using the 3′-OH of the terminal G414 as the nucleophile to attack the splice site between the exons (Figure 2A) (*15*, *16*). This reaction transfers the 3′-exon onto the 3′-end of the intron, which radiolabels the active molecules in the population if the oligonucleotide substrate is 3′-end-labeled. Following the ligation reaction, the RNAs are digested with iodine and the cleavage products resolved by PAGE (Figure 4A). The intensities of individual bands in the m^2GαS and GαS cleavage ladders were compared to the 5′-end-labeled controls to identify the sites of interference. Although the majority of the 107 Gs within the ribozyme were informative, complete data could not be obtained for 10 sites due to the inability to separate the cleavage products at the positions furthest from the location of the radiolabel (G22−G32 for the 3′-exon ligation reaction, and G405−G414 for the 5′-end-labeled control). As might be expected, most positions did not show any effect upon analogue substitution. Greater than 95% of the interference κ values were between 0.67 and 1.5.

Two sites, G212 and G303, showed complete interference with m^2GαS in the 3′-exon ligation reaction (Figure 4A, lanes 8 and 9; Figure 5). Both of these positions are completely conserved among the 131 examples of IC1 and IC2 introns, of which the *Tetrahymena* intron is a member (*33*). The only other Gs that are conserved to this level among IC1

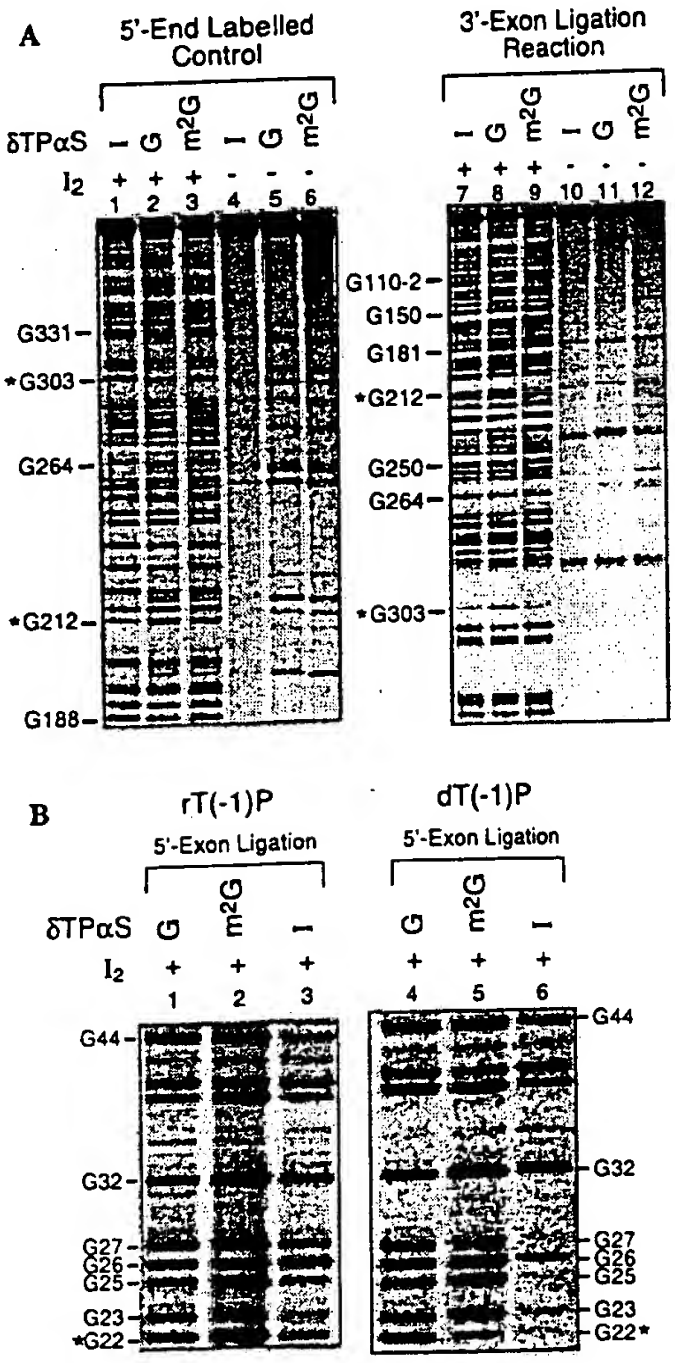


FIGURE 4: Analogue incorporation and interference reactions for 3′- and 5′-exon ligation. (A) Interference mapping for the 3′-exon ligation reaction. The phosphorothioate-tagged analogue incorporated into each RNA is listed above the lane numbers. The nucleotide number corresponding to several of the bands is marked to the left of each gel. The addition (lanes 1−3, 7−9) or omission (lanes 4−6, 10−12) of iodine is indicated. 5′-End Labeled Control: The L-21 G414 5′-end-labeled control showing the extent and positions of analogue incorporation throughout the intron. Cleavage products at positions G212 and G303 (asterisks) show $m^2G\alpha S$ was incorporated at these sites (lane 3). This particular gel was electrophoresed at 75 W for 2.5 h. Longer electrophoretic times were used to improve the signal resolution of the nucleotides toward the 3′-end of the RNA (not shown). 3′-Exon Ligation Reaction: The 3′-exon ligation reaction of L-21 G414 RNA with dT(-1)S. This autoradiogram reveals the sites of analogue interference throughout the intron. The $m^2G\alpha S$ cleavage products at G303 and G212 (asterisks) are of substantially lower intensity than in the 5′-end-labeled control (lane 9). This particular gel was electrophoresed at 75 W for 1.75 h. Longer electrophoretic times were used to resolve the cleavage products toward the 5′-end of the intron (not shown). (B) The 5′-exon ligation reaction of L+1 *Sca*I RNA with rT(-1)P or dT(-1)P in 10 or 4 mM $MgCl_2$, respectively. This autoradiogram shows the interference results for nucleotide positions within the IGS. The asterisk is to call attention to the $m^2G\alpha S$ interference seen at G22. $m^2G\alpha S$ interference was also observed at G212 (not shown). Interference was observed with IαS at all the positions within the IGS (compare lanes 3 and 6). The no-iodine control lanes are omitted, but they were equivalent to those shown in panel A.

and IC2 introns are G22 and G264. G22 was not informative in this assay. G264 is known to be essential to intron function because it forms part of the G binding site, but it does this using functional groups in the major groove of the P7 helix that would not be affected by $m^2G\alpha S$ substitution (*34*). No other sites showed detrimental effects due to $m^2G\alpha S$ substitution, including G111 and G112 where modest interference was previously observed with IαS substitution (*8*). G303 also showed interference with IαS, though there was no IαS interference at G212.

One region of particular interest within the ribozyme is the internal guide sequence (IGS) strand of the P1 helix (Figure 1). Unfortunately, the five Gs within the IGS were uninformative because the cleavage products could not be resolved in the 3′-exon ligation reaction. To gain information about these nucleotides, we used the L+1 *Sca*I ribozyme, which lacks the last 3 nucleotides at the 3′-end of the intron (including G414), but includes the first 21 nucleotides at the 5′-end (*8*). This form of the intron includes a G as the first base of the intron that is equivalent to the exogenous G added to the 5′-end of the intron after the first step of splicing (*35*). In the presence of a 5′-exon oligonucleotide substrate, this intron performs a reaction that is analogous to the reverse of the first step of splicing, wherein it transfers the 5′-exon onto the 5′-end of the intron with concomitant release of the terminal G (*13, 14*) (Figure 2B). This reaction places the radiolabel at the 5′-end of the intron where the IGS nucleotides can be readily resolved. Using this reaction, all five of the Gs within the P1 helix showed interference with IαS, though only G22 is conserved (*8, 19*).

In contrast to the results with IαS interference, G22 was the only position in the IGS that showed interference from $m^2G\alpha S$ substitution (Figure 4B). No $m^2G\alpha S$ interference was detected at G23, G25, G26, or G27. G22 is universally conserved among all known group I introns and is essential for defining the 5′-splice site (*33, 36−38*). In the 5′-exon ligation reaction, strong interference was also detected at G212, though there was no interference at G303 (data not shown). Thus, using two G analogues that alter the exocyclic amine, there are sites where interference was observed with both analogues (G22 and G303), as well as sites where only one of the analogues interfered with activity; G212 only showed interference with $m^2G\alpha S$, while G23, G25, G26, G27, G111, and G112 only showed interference with IαS.

Mutagenesis of the Closing Base Pair in the P4 Helix. Given the reduced stabilities of duplexes containing inosine compared to m^2G (*39−41*), the lack of $m^2G\alpha S$ interference at some of the IαS interference sites suggests that reduced duplex stability is the likely cause of IαS interference at these positions. This interpretation implies that the stability of the closing G−C base pair in the P4 helix might be important for 3′-exon ligation activity. To test this possibility directly in the L-21 G414 ribozyme, we mutated the G112−C208 pair to an A112−U208 pair and measured the rates of 3′-exon ligation with the dT(-1)S substrate under the same conditions as those used for the interference experiments (*16*). k_{cat} for 3′-exon ligation was unaffected by the base pair mutation (0.015 ± 0.002 and 0.014 ± 0.001 min^{-1} for G112−C208 and A112−U208, respectively), but K_m increased by approximately 2.5-fold (54 and 140 nM, respectively).

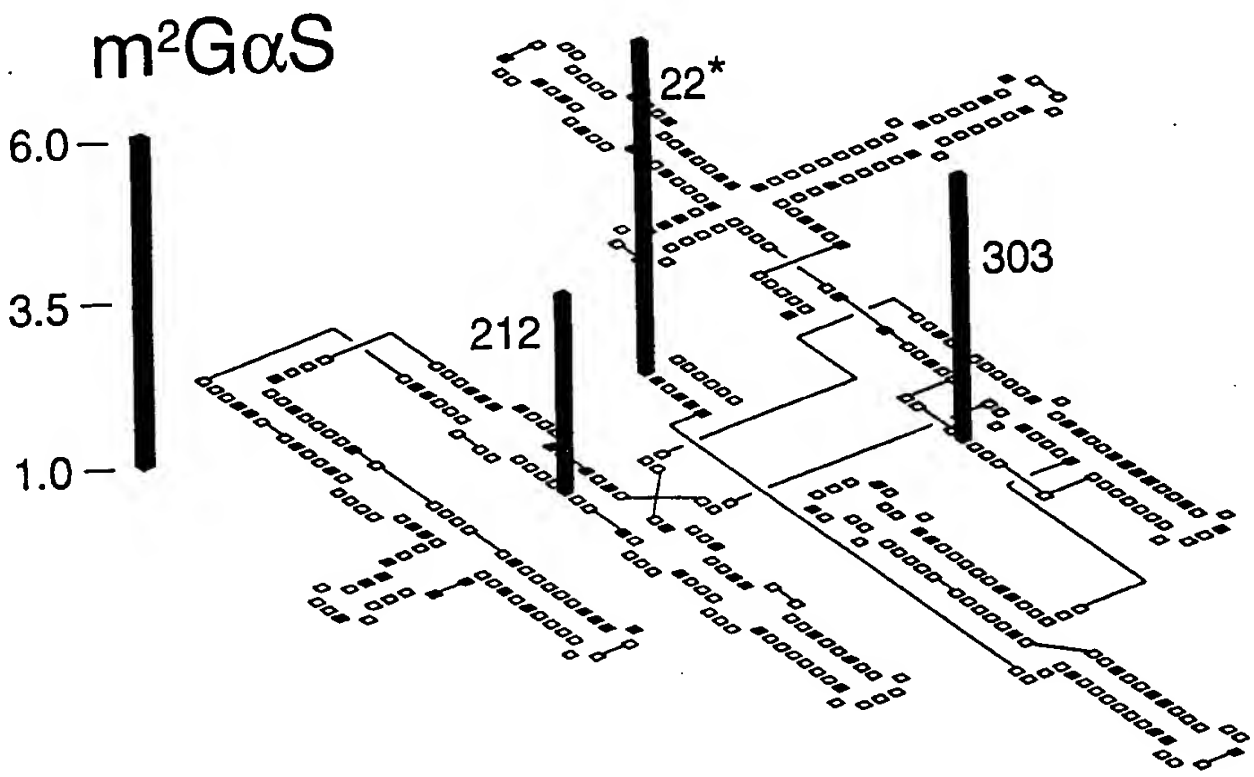


FIGURE 5: Individual histograms for both of the G analogues plotting the magnitude of the interference κ value versus nucleotide position superimposed on the intron secondary structure. Interference κ values ≥ 2.0 are shown as gray bars with the nucleotide number listed adjacent to the bar. Values greater than 6 are assigned a magnitude of 6 for this graph. Black boxes indicate G sites where the κ value was less than 2 (meaning no interference), and white boxes mark nucleotide positions other than G. Gray boxes indicate G sites that were not informative in the assay due to incomplete resolution of the cleavage products on the sequencing gel (G400−G414). The 5′-exon ligation reaction with L+1 *Sca*I RNA was used to obtain values within the IGS (marked with an asterisk). All other values in the histogram were for the 3′-exon ligation reaction using the L-21 G414 RNA. The error in the κ value at each position is $\leq 20\%$. Plotted values are the average of at least two, and as many as eight, independent experimental measurements.

DISCUSSION

Nucleotide analogue interference mapping (NAIM) is a general biochemical method to identify the chemical groups that are important for RNA function (Figure 2C). In a previous report, we used IαS to identify the G positions in the *Tetrahymena* intron where the exocyclic amine is important for 3′-exon and 5′-exon ligation (*8*). These included all five of the Gs within the internal guide sequence (IGS), the closing two G−C base pairs at the top of the P4 helix, and G303 which is located within the J8/7 single-stranded segment that threads through the center of the catalytic core (Figure 5). Based upon sequence conservation and mutagenesis experiments, not all the sites of IαS interference are expected to be involved in tertiary contacts within the ribozyme fold (*33*, *42*).

Inosine substitution, which deletes the N2 exocyclic amine of G, can affect at least two aspects of RNA folding, duplex

stability and tertiary structure formation. Thermodynamic characterization of model RNA duplexes containing inosine substitutions shows that an I−C pair is between 1 and 2 kcal·mol^{-1} less stable than a G−C pair (*39, 40*). Yet, while inosine substitution can destabilize RNA duplexes, it can also have dramatic effects on ribozyme activity if the N2 amine participates in tertiary hydrogen bonding, such as seen at position G22 within the *Tetrahymena* intron (*43*). Because inosine affects both secondary and tertiary stability, it is not possible with IαS interference mapping alone to distinguish between these two contributions to RNA structure.

N^2-Methylguanosine (m^2G) is a second analogue that modifies the exocyclic amine of G, but instead of deleting the functional group, it substitutes one proton of the amine with a methyl group (Figure 3). Duplex studies with m^2G have shown that it can form base pairs with C, U, and A that are at least as stable as those formed by G, which indicates that the methyl group is equally stable on either the s-trans or the s-cis face of the base (*41, 44*). We reasoned that this property might make it possible to use the phosphorothioate of m^2G to probe specifically for essential minor groove tertiary contacts in RNA without encountering the secondary structural effects that are observed with IαS.

Toward this objective, we synthesized the phosphorothioate of N^2-methylguanosine ($m^2G\alpha S$) and randomly incorporated it into two reverse splicing forms of the *Tetrahymena* group I intron using a mutant form of the T7 RNA polymerase. We observed three sites of $m^2G\alpha S$ interference within the *Tetrahymena* intron, G22, G212, and G303 (Figure 5). These sites are located in the P1 helix, P4 helix, and J8/7 single-stranded region, respectively (Figure 1). All three of the Gs are completely conserved among the IC1−IC2 introns (which includes the *Tetrahymena* intron) (*33*), and all three Gs are known to participate in minor groove tertiary hydrogen bonding between helical elements of the ribozyme structure (*2, 3, 45*) (Figure 6).

Interference of P1 Helix Docking into J4/5. G22 forms a wobble pair with U-1 at the 5′-splice site of the intron (*46*), and these two bases are the only conserved residues within the P1 helix (Figure 1) (*19, 42*). While U-1 is essential for holding the G in a wobble configuration, U-1 does not participate directly in ground-state tertiary interactions with the intron core (*43*). Instead, the G utilizes its amine and its 2′-OH to dock into a wobble receptor located within the J4/5 region (*3, 47, 48*). Within a G·U wobble pair, neither of the amino protons participate in duplex formation; however, both protons of the G22 amino group participate in tertiary hydrogen bonds with the two consecutively stacked sheared A·A pairs that constitute the wobble receptor (*3*). One proton donates a hydrogen bond to the N3 of A207 while the second proton donates a hydrogen bond to the 2′-OH of A207 (Figure 6A). Interference with $m^2G\alpha S$ and IαS is in full agreement with this model for P1 helix docking, because placement of the $m^2G\alpha S$ methyl group onto either the s-cis or the s-trans face of the nucleotide would disrupt a hydrogen bond essential for docking the 5′-exon into the active site.

Interference of the P4 Interaction with the A-Rich Bulge. G212 base-pairs with C109 in the P4 helix. A G−C pair utilizes one of the exocyclic amino protons for duplex formation, but leaves the second proton unpaired in the minor groove. The crystal structure of the P4−P6 domain demonstrated that G212, and nucleotides surrounding it, makes extensive minor groove tertiary interactions with the A-rich bulge of the P5abc extension (Figure 1) (*2, 49*). These contacts are largely sequence-nonspecific hydrogen bonds to 2′-OHs on the P4 helix. The single exception is a tertiary hydrogen bond between the N2 amine of G212 and the N3 of A184 (Figure 6B). Strong interference at G212 with $m^2G\alpha S$ suggests that the interaction between P4 and the A-rich bulge is important for activity, though the absence of IαS interference indicates that deletion of the hydrogen bond is not sufficient to affect intron splicing. In this closely packed region of the RNA, the additional steric bulk of the N2-methyl group in the P4 minor groove is likely to be more destabilizing than deletion of the amine.

$m^2G\alpha S$ interference at G212 is the first example of significant loss of intron splicing activity resulting from a

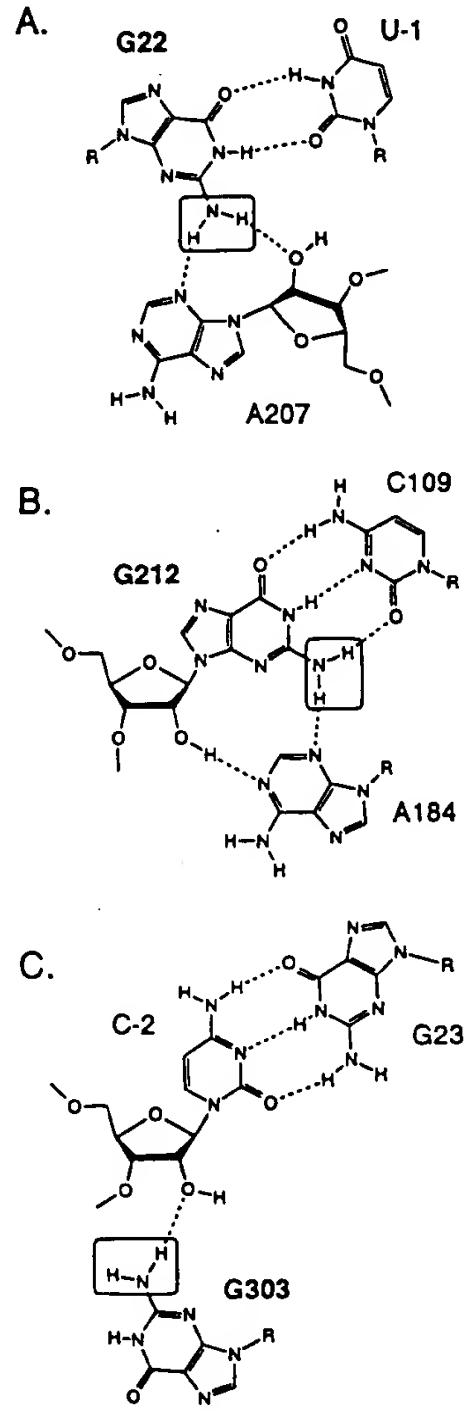


FIGURE 6: Tertiary hydrogen bonding interactions postulated for the three Gs (bold) that show $m^2G\alpha S$ interference. In each case, the amino group that shows interference is highlighted with a box. (A) G22 forms a wobble pair with U-1 and makes a tertiary interaction with A207 (*3*). (B) G212 forms a Watson−Crick pair with C109 and makes a tertiary interaction with A184 (*2*). (C) G303 is in a single-stranded region of the ribozyme, but makes a tertiary interaction with C-2 (*45*).

single point mutation or functional group change in the interface between P5abc and the P4−P6 helical stack (*9*, *50*). Apparently, there is sufficient energetic redundancy to overcome most minor structural alterations in the context of the intact intron (*50*). The P4−P6 crystal structure suggests three other sites within the domain that are strong candidates for $m^2G\alpha S$ interference (*2*). The N2 amino groups of G150 and G250 participate in the tetraloop−tetraloop receptor interaction between P5b and P6a, and the amine of G181 helps form the substructure of the A-rich bulge; however, interference was not observed at any of these three sites in either ligation assay (Figure 5). $m^2G\alpha S$ interference at G212, but not at G150 or G250, suggests that the A-rich bulge interaction with P4 may be more important for intron function than the tetraloop/tetraloop receptor interaction.

Interference from P1 Helix Docking into J8/7. Unlike the other two sites that demonstrated $m^2G\alpha S$ interference, G303 is in a single-stranded segment of the intron, J8/7 (Figure 1), which forms an extended triple helix complex with the minor groove of the P1 helix (*45*). Interference suppression experiments using a substrate with a 2′-deoxy substitution at C-2 demonstrated that G303 forms a tertiary hydrogen bond between its exocyclic amine and the 2′-OH of C-2 in the P1 helix (Figures 1 and 6C) (*45*). However, the interaction proposed to form between these two residues only utilizes one of the G303 amino protons. Presumably $m^2G\alpha S$ could adopt the s-cis rotamer to facilitate hydrogen bonding by the s-trans proton of the amine. The fact that interference is observed with both $I\alpha S$ and $m^2G\alpha S$ suggests either that the s-cis proton is involved in an additional hydrogen bond or that there is insufficient space in the P1−J8/7 packing interface to accommodate an s-cis methyl group.

It is somewhat curious that $I\alpha S$ and $m^2G\alpha S$ interference is observed at G303 for the 3′-exon but not the 5′-exon ligation reaction. While this could reflect an important conformational difference between the first and second steps of splicing, the more likely explanation is that there is additional energetic redundancy for P1 helix docking within the 5′-exon ligation reaction. In the L+1 *Sca*I construct, the reactive G is covalently connected to the P1 helix. Additional stabilization energy for P1 docking is provided by G binding into the P7 helix, which would make the G303 interaction with C-2 less critical. In the L-21 G414 ribozyme, the reactive G is at the 3′-end of the RNA where it does not contribute directly to P1 helix binding. In this arrangement, the interaction between G303 and C-2 is necessary for alignment of the substrate into the active site.

Interference with $I\alpha S$ but Not $m^2G\alpha S$ Suggests Duplex Stability Is Important. There are six sites that showed interference with $I\alpha S$ that were not sensitive to $m^2G\alpha S$ substitution. All the sites are in double-stranded regions of the intron, and they cluster into two helices. Four of the sites (G23, G25, G26, and G27) are in the P1 helix, and two of the sites (G111 and G112) are at the top of the P4 helix just below the sheared A·A pairs in J4/5.

The only phylogenetically conserved sequence in the P1 helix is the G22·U-1 base pair at the cleavage site. The remaining positions maintain complementarity between the 5′-exon and the IGS, but the primary sequence is not conserved and can be mutated as long as base-pairing is retained (*42*). This lack of sequence conservation suggests that base-specific functional groups do not participate directly in tertiary interactions with the ribozyme catalytic core, which implies that $I\alpha S$ interference at these sites results from reduced secondary structural stability rather than the loss of a tertiary hydrogen bond to an N2-exocyclic amine. In a competitive environment where multiple ribozymes (50 nM) are vying for a slow reacting substrate (k_{cat} = 0.015 min^{-1}, k_{off} > 0.1 min^{-1}) that is present in limiting quantity (10 nM, 0.2 equiv), the inability of some ribozymes to form a stable duplex between the IGS and the substrate confers a significant disadvantage compared to other variants in the population. In contrast, interference was not observed with $m^2G\alpha S$ at the nonconserved positions in the P1 helix, because this analogue does not affect duplex stability (*41*). Therefore, interference from $I\alpha S$, but not from $m^2G\alpha S$, defines a biochemical signature to identify positions where secondary structural stability is critical for RNA function.

This pattern of interference is also seen within the P4 helix where nucleotides G111 and G112 showed weak interference with $I\alpha S$. This is a challenging region of the sequence to analyze experimentally by NAIM because it is difficult to resolve the cleavage products for the 3 consecutive Gs (G110−G112) that are located more than 300 nucleotides from the radiolabel at the 3′-terminus. Furthermore, interference at these sites appears to be highly dependent upon the reaction conditions. For example, $I\alpha S$ interference at G111 and G112 was not observed under slightly modified divalent metal conditions (such as 3 mM Mg^{2+} and 1 mM Mn^{2+}; M. Kronman and S. A. Strobel, unpublished results) nor was it observed in the 5′-exon ligation reaction (*8*). Nevertheless, the interference data with $I\alpha S$ and $m^2G\alpha S$ at 4 mM Mg^{2+} imply that ribozymes with a stable G112−C208 closing base pair in the P4 helix have a modest selective advantage over those with an $I\alpha S$112−C208 pair.

To directly test the importance of the P4 closing base pair, we mutated the G112−C208 pair to A−U in the context of the L-21 G414 ribozyme and measured the second-order rate constant for 3′-exon ligation under the same conditions as those used for the interference experiments (*16*). k_{cat} for the mutant enzyme was the same as the wild-type, but the K_m was 2.5-fold higher. The magnitude of the $I\alpha S$ interference suggested that the effect would be larger, but it is possible that the A−U mutation is not as destabilizing as the $I\alpha S$−C pair that occurs in the interference experiment. Thermodynamic measurements of model duplexes have shown that an I−C pair is in fact slightly less stable (0.3 kcal·mol^{-1}) than an A−U pair (*39*).

The G112−C208 pair is immediately below two consecutive sheared A·A pairs that act as a wobble receptor for the P1 helix (Figure 1) (*3*). Thermodynamic measurements of model duplexes containing consecutive A·A or G·A mismatches have demonstrated that the mispairs show a net stabilization from a G−C closing pair, but a net destabilization from an A−U closing pair (*51*, *52*). One dramatic consequence of this destabilization is that mutation of the closing base pair converts the G·A mispair from a sheared conformation to a imino hydrogen-bonded conformation (*53*, *54*). While the A·A pair cannot undergo this secondary structural transition, there is a precedent to argue that the stability of the P4 closing pair is important to stabilize the consecutively stacked sheared A·A pairs located immediately above it.

m^2GαS Interference Only at Sites of Direct Tertiary Hydrogen Bonding. Immediately adjacent to the sites of m^2GαS interference at G22 and G212 within the P1 and P4 helices, respectively, there are G−C pairs that do not show interference (Figures 1 and 5). These Gs participate in tertiary hydrogen bonds, but utilize functional groups other than their N2 amines. For example the 2′-OHs of the G23−C-2 pair each are thought to participate in a direct hydrogen bond; the 2′-OH of G23 bonds with the 2′-OH of C208 (*9*), and the 2′-OH of C-2 bonds with the N2 amine of G303 (*45*). Despite tremendous close packing within the minor groove face of this region of the molecule, interference was not observed with m^2GαS at G23. A similar pattern was observed in the P4 helix, where the 2′-OHs of G110, G212, and C109 all participate in hydrogen bonds to groups within the A-rich bulge. Nevertheless, m^2GαS interference was not observed in the G110−C211 base pair above the C109−G212 pair, nor was it observed in the G108−C213 base pair below G212. Thus, within the *Tetrahymena* ribozyme, m^2GαS interference is only observed at Gs that participate directly in tertiary hydrogen bonds via their exocyclic amines, not just at Gs that are closely packed within the tertiary structure.

All three of the sites of m^2GαS interference within the *Tetrahymena* intron are highly conserved and participate in tertiary hydrogen bonding interactions via their N2 amino groups. The sites of interference within this intron occurred at a G−C pair, a G·U pair, and an unpaired G. Given that the wide and shallow RNA minor groove is likely to be a common interface for RNA helix packing, and that there are relatively few reagents that probe this face of the helix, m^2GαS provides a valuable tool for RNA structure/function mapping in less well-characterized RNAs. This is particularly true if m^2GαS is used in combination with IαS, because together these analogues make it possible to differentiate sites of essential secondary structural stability from sites of tertiary hydrogen bonding.

ACKNOWLEDGMENT

We thank Charles Cheng for assistance with nucleotide synthesis, Rui Sousa for the gift of the Y639F mutant T7 polymerase clone, and Alexander Szewczak, Sean P. Ryder, and Juliane Strauss-Soukkup for critical comments on the manuscript.

REFERENCES

1. Strobel, S. A., and Doudna, J. A. (1997) *Trends Biochem. Sci. 22*, 262−266.
2. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) *Science 273*, 1678−1685.
3. Strobel, S. A., Ortoleva-Donnelly, L., Ryder, S. P., Cate, J. H., and Moncoeur, E. (1998) *Nat. Struct. Biol. 5*, 60−66.
4. Litt, M., and Hancock, V. (1967) *Biochemistry 6*, 1848−1854.
5. Noller, H. F. (1974) *Biochemistry 13*, 4694−4703.
6. Gaur, R. K., and Krupp, G. (1993) *Nucleic Acids Res. 21*, 21−26.
7. Hardt, W. D., Erdmann, V. A., and Hartmann, R. K. (1996) *RNA 2*, 1189−1198.
8. Strobel, S. A., and Shetty, K. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 2903−2908.
9. Ortoleva-Donnelly, L., Szewczak, A. A., Gutell, R. R., and Strobel, S. A. (1998) *RNA 4*, 498−519.
10. Gish, G., and Eckstein, F. (1988) *Science 240*, 1520−1522.
11. Conrad, F., Hanne, A., Gaur, R. K., and Krupp, G. (1995) *Nucleic Acids Res. 23*, 1845−1853.
12. Cech, T. R. (1990) *Annu. Rev. Biochem. 59*, 543−568.
13. Woodson, S. A., and Cech, T. R. (1989) *Cell 57*, 335−345.
14. Green, R., Ellington, A. D., and Szostak, J. W. (1990) *Nature 347*, 406−408.
15. Beaudry, A. A., and Joyce, G. F. (1992) *Science 257*, 635−641.
16. Mei, R., and Herschlag, D. (1996) *Biochemistry 35*, 5796−5809.
17. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) *Science 273*, 1696−1699.
18. Cate, J. H., Hanna, R. L., and Doudna, J. A. (1997) *Nat. Struct. Biol. 4*, 553−558.
19. Michel, F., and Westhof, E. (1990) *J. Mol. Biol. 216*, 585−610.
20. Lehnert, V., Jaeger, L., Michel, F., and Westhof, E. (1996) *Chem. Biol. 3*, 993−1009.
21. Robins, M. J., and Uznanski, B. (1981) *Can. J. Chem. 59*, 2608−2611.
22. Avino, A. M., Mayordomo, A., Espuny, R., Bach, M., and Eritja, R. (1995) *Nucleotides Nucleosides 14*, 1613−1617.
23. Rife, J., Cheng, C., Moore, P. B., and Strobel, S. A. (1998) *Nucleotides Nucleosides* (in press).
24. Sekine, M., and Satoh, T. (1991) *J. Org. Chem. 56*, 1224−1227.
25. Arabshahi, A., and Frey, P. A. (1994) *Biochem. Biophys. Res. Commun. 204*, 150−155.
26. Sousa, R., and Padilla, R. (1995) *EMBO J. 14*, 4609−4621.
27. Christian, E. L., and Yarus, M. (1992) *J. Mol. Biol. 228*, 743−758.
28. Lingner, J., and Keller, W. (1993) *Nucleic Acids Res. 21*, 2917−2920.
29. Steitz, T. A., Beese, L., Freemont, P. S., Friedman, J. M., and Sanderson, M. R. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 465−471.
30. Huang, Y., Eckstein, F., Padilla, R., and Sousa, R. (1997) *Biochemistry 36*, 8231−8242.
31. Huang, Y., Beaudry, A., McSwiggen, J., and Sousa, R. (1997) *Biochemistry 36*, 13718−13728.
32. Raines, K., and Gottlieb, P. A. (1998) *RNA 4*, 340−345.
33. Damberger, S. H., and Gutell, R. R. (1994) *Nucleic Acids Res. 22*, 3508−3510.
34. Michel, F., Hanna, M., Green, R., Bartel, D. P., and Szostak, J. W. (1989) *Nature 342*, 391−395.
35. Grabowski, P. J., Zaug, A. J., and Cech, T. R. (1981) *Cell 23*, 467−476.
36. Doudna, J. A., Cormack, B. P., and Szostak, J. W. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7402−7406.
37. Barfod, E. T., and Cech, T. R. (1989) *Mol. Cell. Biol. 9*, 3657−3666.
38. Strobel, S. A., and Cech, T. R. (1996) *Biochemistry 35*, 1201−1211.
39. Turner, D. H., Sugimoto, N., Kierzek, R., and Dreiker, S. D. (1987) *J. Am. Chem. Soc. 109*, 3783−3785.
40. Strobel, S. A., Cech, T. R., Usman, N., and Beigelman, L. (1994) *Biochemistry 33*, 13824−13853.
41. Rife, J., Cheng, C., Moore, P. B., and Strobel, S. A. (1998) *Nucleic Acids Res.*, In press.
42. Murphy, F. L., and Cech, T. R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9218−9222.
43. Strobel, S. A., and Cech, T. R. (1995) *Science 267*, 675−679.
44. Engel, J. D., and von Hippel, P. H. (1974) *Biochemistry 13*, 4143−4158.
45. Szewczak, A. A., Ortoleva-Donnelly, L., Ryder, S. P., and Strobel, S. A. (1998) (submitted for publication).
46. Waring, R. B., Towner, P., Minter, S. J., and Davies, R. W. (1986) *Nature 321*, 133−139.

47. Strobel, S. A., and Cech, T. R. (1993) *Biochemistry 32*, 13593−13604.
48. Wang, J.-F., Downs, W. D., and Cech, T. R. (1993) *Science 260*, 504−508.
49. Flor, P. J., Flanegan, J. B., and Cech, T. R. (1989) *EMBO J. 8*, 3391−3399.
50. Laggerbauer, B., Murphy, F. L., and Cech, T. R. (1994) *EMBO J. 13*, 2669−2676.
51. Walter, A. E., Wu, M., and Turner, D. H. (1994) *Biochemistry 33*, 11349−11354.
52. Wu, M., McDowell, J. A., and Turner, D. H. (1995) *Biochemistry 34*, 3204−3211.
53. SantaLucia, J., and Turner, D. H. (1993) *Biochemistry 32*, 12612−12623.
54. Wu, M., and Turner, D. H. (1996) *Biochemistry 35*, 9677−9689.

BI980723J